| FORM PTO-1390 (REV 11-2000) | U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE | ATTORNEY'S DOCKET NUMBER 35-213 |
|---|---|---|

## TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

U S APPLICATION NO. (If known, see 37 C F R 1 5)
**09/913159**
Unknown

| INTERNATIONAL APPLICATION NO. PCT/EP99/09759 | INTERNATIONAL FILING DATE 3 December 1999 | PRIORITY DATE CLAIMED 12 February 1999 |
|---|---|---|

TITLE OF INVENTION

**VIRAL VACCINE**

APPLICANT(S) FOR DO/EO/US

**SCHREIBER**

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.

2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.

3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.

4. ☒ The U.S. has been elected by the expiration of 19 months from the priority date (Article 31).

5. A copy of the International Application as filed (35 U.S.C. 371(c)(2)).

   a. ☐ is attached hereto (required only if not communicated by the International Bureau).

   b. ☒ has been communicated by the International Bureau.

   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).

6. ☒ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).

   a. ☒ is attached hereto.

   b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).

7. ☒ Amendments to the claims of the International Application under <u>PCT Article 34</u>

   a. ☐ are attached hereto (required only if not communicated by the International Bureau).

   b. ☒ have been communicated by the International Bureau.

   c. ☐ have not been made; however, the time limit for making such amendments has **NOT** expired.

   d. ☐ have not been made and will not be made.

8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).

9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).

10. ☒ An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11 To 20 below concern document(s) or information included:**

11. ☐ An Information Disclosure Statement under 37 C.F.R. 1.97 and 1.98.

12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 C.F.R. 3.28 and 3.31 is included.

13. ☒ A FIRST preliminary amendment.

14. ☐ A SECOND or SUBSEQUENT preliminary amendment.

15. ☐ A substitute specification.

16. ☐ A change of power of attorney and/or address letter.

17. ☒ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821-1.825.

18. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).

19. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).

20. ☒ Other items or information. PTO-1449/ International Search Report/ Sequence Listing/ Front page of the PCT Publication/ Certification of Translation/Letter re: Sequence Listing
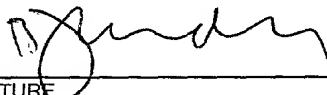
545026

| U.S. APPLICATION NO. *(If known see 37 C.F.R 1.5)* | INTERNATIONAL APPLICATION NO. PCT/EP99/09759 | ATTORNEY'S DOCKET NUMBER 35-213 |
|---|---|---|
| Unknown 09/913159 | | |

**21.** ☒ The following fees are submitted:

**BASIC NATIONAL FEE (37 C.F.R. 1.492(a)(1)-(5):**

| CALCULATIONS | PTO USE ONLY |
|---|---|

-- Neither international preliminary examination fee (37 C.F.R. 1.482) nor international search fee (37 C.F.R. 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO .......................$1000.00

-- International preliminary examination fee (37 C.F.R. 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO...................$860.00

-- International preliminary examination fee (37 C.F.R. 1.482) not paid to USPTO but international search fee (37 C.F.R. 1.445(a)(2)) paid to USPTO.....................$710.00

-- International preliminary examination fee (37 C.F.R. 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4)..............................$690.00

-- International preliminary examination fee (37 C.F.R. 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4)....................................$100.00

| | | | | | |
|---|---|---|---|---|---|
| **ENTER APPROPRIATE BASIC FEE AMOUNT =** | | | | $ | 860.00 |

Surcharge of $130.00 for furnishing the oath or declaration later than ☐ 20  ☒ 30 months from the earliest claimed priority date (37 C.F.R. 1.492(e)).

| | | $ | 130.00 |
|---|---|---|---|

| CLAIMS | NUMBER FILED | NUMBER EXTRA | RATE | | |
|---|---|---|---|---|---|
| Total Claims | 54 | -20 = | 34 | X | $18.00 |
| Independent Claims | 13 | -3 = | 10 | X | $80.00 |
| MULTIPLE DEPENDENT CLAIMS(S) (if applicable) | | | | $270.00 | |

| | |
|---|---|
| $ 612.00 | |
| 800.00 | |
| $ 0.00 | |
| **TOTAL OF ABOVE CALCULATIONS =** $ 2402.00 | |

☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.

| | 0.00 |
|---|---|
| **SUBTOTAL =** | $ 2402.00 |

Processing fee of $130.00, for furnishing the English Translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 C.F.R. 1.492(f)).  +

| | 0.00 |
|---|---|
| **TOTAL NATIONAL FEE =** | $ 2402.00 |

Fee for recording the enclosed assignment (37 C.F.R. 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 C.F.R. 3.28, 3.31). **$40.00** per property  +

| | $ 0.00 |
|---|---|
| Fee for Petition to Revive Unintentionally Abandoned Application ($1240.00 – Small Entity = $620.00) | $ 0.00 |
| **TOTAL FEES ENCLOSED =** | $ 2402.00 |

| Amount to be: | | |
|---|---|---|
| refunded | $ | |
| Charged | $ | |

a. ☒ A check in the amount of $2402.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. 14-1140 in the amount of $_____ to cover the above fees. A duplicate copy of this form is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 14-1140. A duplicate copy of this form is enclosed.

d. ☒ The entire content of the foreign application(s), referred to in this application is/are hereby incorporated by reference in this application.

**NOTE: Where an appropriate time limit under 37 C.F.R. 1.494 or 1.495 has not been met, a petition to revive (37 C.F.R. 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

NIXON & VANDERHYE P.C.
1100 North Glebe Road, 8th Floor
Arlington, Virginia 22201-4714
Telephone: (703) 816-4000

SIGNATURE

**B. J. Sadoff**
NAME

**36,663**                    **August 10, 2001**
REGISTRATION NUMBER     Date

545026

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

**SCHREIBER**

Serial No. **09/913,159**

Filed: **August 10, 2001**

For: **VIRAL VACCINE**

Atty. Ref.: **35-213**

Group: **Unknown**

Examiner: **Unknown**

\* \* \* \* \* \* \* \* \* \* \*
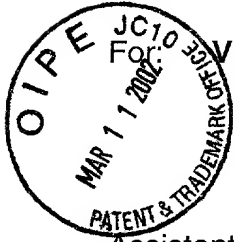
**March 11, 2002**

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

## AMENDMENT

Responsive to the Notification dated January 18, 2002, and further to the

Request of February 6, 2002, entry and consideration of the following amendments and

remarks are requested.

## IN THE SPECIFICATION

Amend the specification as follows:

Insert the attached Sequence Listing in place of the copy of the same filed

August 10, 2001.

## REMARKS

Reconsideration is requested.

609728

A copy of the Notification of Missing Requirements dated January 18, 2002, and

Raw Sequence Listing Error Report attached to the same, are attached. The

specification has been amended to include the attached Sequence Listing.

The attached paper and computer-readable copies of the Sequence Listing are

the same. No new matter has been added. A separate Statement to this effect is

attached.
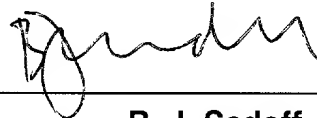
Nothing further is believed to be required in response to the Notification of

Missing Requirements dated January 18, 2002, however the Office is requested to

contact the undersigned if otherwise.

An early and favorable Action on the merits is requested.

Respectfully submitted,

**NIXON & VANDERHYE P.C.**

By: _____

**B. J. Sadoff**
Reg. No. **36,663**

**BJS:eaw**
1100 North Glebe Road, 8th Floor
Arlington, VA 22201-4714
Telephone: (703) 816-4000
Facsimile: (703) 816-4100

#7

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

**SCHREIBER**

Serial No.  **09/913,159**

Filed:  **August 10, 2001**

For  **VIRAL VACCINE**

Atty. Ref.:  **35-213**

Group:  **Unknown**

Examiner:  **Unknown**

\* \* \* \* \* \* \* \* \* \*

**March 11, 2002**

Assistant Commissioner for Patents
Washington, DC  20231

Sir:

## STATEMENT

The attached paper and computer-readable copies of the Sequence Listing are

the same.  No new matter has been added.

Respectfully submitted,

**NIXON & VANDERHYE P.C.**

By: _____

**B. J. Sadoff**
Reg. No. **36,663**

**BJS:eaw**
1100 North Glebe Road, 8th Floor
Arlington, VA 22201-4714
Telephone: (703) 816-4000
Facsimile:  (703) 816-4100

609729

RAW SEQUENCE LISTING
PATENT APPLICATION:   US/09/913,159A

DATE: 04/29/2002
TIME: 11:20:40

Input Set : **A:\seq listing.txt**
Output Set: **N:\CRF3\04292002\I913159A.raw**

```
 4 <110> APPLICANT: Strathmann AG & Co.
 6 <120> TITLE OF INVENTION: Virus-Vaccine
 8 <130> FILE REFERENCE: P057760
C--> 10 <140> CURRENT APPLICATION NUMBER: US/09/913,159A
C--> 11 <141> CURRENT FILING DATE: 2002-03-11
13 <150> PRIOR APPLICATION NUMBER: 199 07 485.2
14 <151> PRIOR FILING DATE: 1999-02-12
16 <160> NUMBER OF SEQ ID NOS: 12
18 <170> SOFTWARE: PatentIn Ver. 2.1
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 9709
22 <212> TYPE: DNA
23 <213> ORGANISM: Human immunodeficiency virus
25 <400> SEQUENCE: 1
26 tggaagggct aatttggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca 60
27 cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac 120
28 tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca 180
29 aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg 240
30 agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccgag 300
31 agctgcatcc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg 360
32 ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga gccctcagat 420
33 gctacatata agcagctgct ttttgcctgt actgggtctc tctggttaga ccagatctga 480
34 gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct 540
35 tgagtgctca agtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600
36 agacccttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacttgaaag 660
37 cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgctgaa gcgcgcacgg 720
38 caagaggcga ggggcggcga ctggtgagta cgccaaaaat tttgactagc ggaggctaga 780
39 aggagagaga tgggtgcgag agcgtcggta ttaagcgggg gagaattaga taaatgggaa 840
40 aaaattcggt taaggccagg gggaagaaa caatataaac taaacatat agtatgggca 900
41 agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaaggctgt 960
42 agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca 1020
43 ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc 1080
44 aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa 1140
45 gcagcagctg acacaggaaa caacagccag gtcagccaaa attaccctat agtgcagaac 1200
46 ctccaggggc aaatggtaca tcaggccata tcacctagaa ctttaaatgc atgggtaaaa 1260
47 gtagtagaag agaaggcttt cagcccagaa gtaataccca tgttttcagc attatcagaa 1320
48 ggagccaccc cacaagattt aaataccatg ctaaacacag tggggggaca tcaagcagcc 1380
49 atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca 1440
50 gtgcatgcag gcctattgc accaggccag atgagagaac aaggggaag tgacatagca 1500
51 ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca 1560
52 gtaggagaaa tctataaaag atggataatc ctgggattaa ataaaatagt aagaatgtat 1620
53 agccctacca gcattctgga cataagacaa ggaccaaagg aaccctttag agactatgta 1680
54 gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg 1740
```

RAW SEQUENCE LISTING                    DATE: 04/29/2002
PATENT APPLICATION:   US/09/913,159A     TIME: 11:20:40

Input Set : **A:\seq listing.txt**
Output Set: **N:\CRF3\04292002\I913159A.raw**

```
 55 acagaaacct tgttggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg 1800
 56 ggaccaggag cgacactaga agaaatgatg acagcatgtc agggagtggg gggacccggc 1860
 57 cataaagcaa gagtttttggc tgaagcaatg agccaagtaa caaatccagc taccataatg 1920
 58 atacagaaag gcaattttag gaaccaaaga aagactgtta agtgtttcaa ttgtggcaaa 1980
 59 gaagggcaca tagccaaaaa ttgcagggcc cctaggaaaa agggctgttg gaaatgtgga 2040
 60 aaggaaggac accaaatgaa agattgtact gagagacagg ctaatttttt agggaagatc 2100
 61 tggccttccc acaagggaag gccagggaat tttcttcaga gcagaccaga gccaacagcc 2160
 62 ccaccagaag agagcttcag gtttggggaa gagacaacaa ctccctctca gaagcaggag 2220
 63 ccgatagaca aggaactgta tcctttagct tccctcagat cactctttgg cagcgacccc 2280
 64 tcgtcacaat aaagataggg gggcaattaa aggaagctct attagataca ggagcagatg 2340
 65 atacagtatt agaagaaatg aatttgccag gaagatggaa accaaaaatg ataggggggaa 2400
 66 ttggaggttt tatcaaagta ggacagtatg atcagatact catagaaatc tgcggacata 2460
 67 aagctatagg tacagtatta gtaggaccta cacctgtcaa cataattgga agaaatctgt 2520
 68 tgactcagat tggctgcact ttaaattttc ccattagtcc tattgagact gtaccagtaa 2580
 69 aattaaagcc aggaatggat ggcccaaaag ttaaacaatg gccattgaca gaagaaaaaa 2640
 70 taaaagcatt agtagaaatt tgtacagaaa tggaaaagga aggaaaaatt tcaaaaattg 2700
 71 ggcctgaaaa tccatacaat actccagtat ttgccataaa gaaaaaagac agtactaaat 2760
 72 ggagaaaaat agtagatttc agagaactta ataagagaac tcaagatttc tgggaagttc 2820
 73 aattaggaat accacatcct gcagggttaa aacagaaaaa atcagtaaca gtactggatg 2880
 74 tgggcgatgc atatttttca gttcccttag ataaagactt caggaagtat actgcattta 2940
 75 ccatacctag tataaacaat gagacaccag ggattagata tcagtacaat gtgcttccac 3000
 76 agggatggaa aggatcacca gcaatattcc agtgtagcat gacaaaaatc ttagagcctt 3060
 77 ttagaaaaca aaatccagac atagtcatct atcaatacat ggatgatttg tatgtaggat 3120
 78 ctgacttaga aataggggcag catagaacaa aaatagagga actgagacaa catctgttga 3180
 79 ggtgggggatt taccacacca gacaaaaaac atcagaaaga acctccattc ctttggatgg 3240
 80 gttatgaact ccatcctgat aaatggacag tacagcctat agtgctgcca gaaaaggaca 3300
 81 gctggactgt caatgacata cagaaattag tgggaaaatt gaattgggca agtcagattt 3360
 82 atgcagggat taaagtaagg caattatgta aacttcttag gggaaccaaa gcactaacag 3420
 83 aagtagtacc actaacagaa gaagcagagc tagaactggc agaaaacagg gagattctaa 3480
 84 aagaaccggt acatggagtg tattatgacc catcaaaaga cttaatagca gaaatacaga 3540
 85 agcaggggca aggccaatgg acatatcaaa tttatcaaga gccatttaaa aatctgaaaa 3600
 86 caggaaaata tgcaagaatg aagggtgccc acactaatga tgtgaaacaa ttaacagagg 3660
 87 cagtacaaaa aatagccaca gaaagcatag taatatgggg aaagactcct aaatttaaat 3720
 88 tacccataca aaaggaaaca tgggaagcat ggtggacaga gtattggcaa gccacctgga 3780
 89 ttcctgagtg ggagtttgtc aatacccctc ccttagtgaa gttatggtac cagttagaga 3840
 90 aagaacccat aataggagca gaaactttct atgtagatgg ggcagccaat agggaaacta 3900
 91 aattaggaaa agcaggatat gtaactgaca gaggaagaca aaaagttgtc cccctaacgg 3960
 92 acacaacaaa tcagaagact gagttacaag caattcatct agctttgcag gattcgggat 4020
 93 tagaagtaaa catagtgaca gactcacaat atgcattggg aatcattcaa gcacaaccag 4080
 94 ataagagtga atcagagtta gtcagtcaaa taatagagca gttaataaaa aaggaaaaag 4140
 95 tctacctggc atgggtacca gcacacaaag gaattggagg aaatgaacaa gtagatgggt 4200
 96 tggtcagtgc tggaatcagg aaagtactat ttttagatgg aatagataag gcccaagaag 4260
 97 aacatgagaa atatcacagt aattggagag caatggctag tgattttaac ctaccacctg 4320
 98 tagtagcaaa agaaatagta gccagctgtg ataaatgtca gctaaaaggg gaagccatgc 4380
 99 atggacaagt agactgtagc ccaggaatat ggcagctaga ttgtacacat ttagaaggaa 4440
100 aagttatctt ggtagcagtt catgtagcca gtggatatat agaagcagaa gtaattccag 4500
101 cagagacagg gcaagaaaca gcatacttcc tcttaaaatt agcaggaaga tggccagtaa 4560
102 aaacagtaca tacagacaat ggcagcaatt tcaccagtac tacagttaag gccgcctgtt 4620
103 ggtgggcggg gatcaagcag gaatttggca ttccctacaa tccccaaagt caaggagtaa 4680
```

file://C:\CRF3\Outhold\VsrI913159A.htm                          4/29/02

RAW SEQUENCE LISTING              DATE: 04/29/2002
PATENT APPLICATION:  US/09/913,159A       TIME: 11:20:40

Input Set : A:\seq listing.txt
Output Set: N:\CRF3\04292002\I913159A.raw

```
104 tagaatctat gaataaagaa ttaaagaaaa ttataggaca ggtaagagat caggctgaac 4740
105 atcttaagac agcagtacaa atggcagtat tcatccacaa ttttaaaaga aaaggggggga 4800
106 ttggggggta cagtgcaggg gaaagaatag tagacataat agcaacagac atacaaacta 4860
107 aagaattaca aaaacaaatt acaaaaattc aaaattttcg ggtttattac agggacagca 4920
108 gagatccagt ttggaaagga ccagcaaagc tcctctggaa aggtgaaggg gcagtagtaa 4980
109 tacaagataa tagtgacata aaagtagtgc caagaagaaa agcaaagatc atcagggatt 5040
110 atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattaacaca 5100
111 tggaaaagat tagtaaaaca ccatatgtat atttcaagga aagctaagga ctggttttat 5160
112 agacatcact atgaaagtac taatccaaaa ataagttcag aagtacacat cccactaggg 5220
113 gatgctaaat tagtaataac aacatattgg ggtctgcata caggagaaag agactggcat 5280
114 ttgggtcagg gagtctccat agaatggagg aaaaagagat atagcacaca agtagaccct 5340
115 gacctagcag accaactaat tcatctgcac tattttgatt gtttttcaga atctgctata 5400
116 agaaatacca tattaggacg tatagttagt cctaggtgtg aatatcaagc aggacataac 5460
117 aaggtaggat ctctacagta cttggcacta gcagcattaa taaaaccaaa acagataaag 5520
118 ccacctttgc ctagtgttag gaaactgaca gaggacagat ggaacaagcc ccagaagacc 5580
119 aagggccaca gagggagcca tacaatgaat ggacactaga gctttttagag gaacttaaga 5640
120 gtgaagctgt tagacatttt cctaggatat ggctccataa cttaggacaa catatctatg 5700
121 aaacttacgg ggatacttgg gcaggagtgg aagccataat aagaattctg caacaactgc 5760
122 tgtttatcca tttcagaatt gggtgtcgac atagcagaat aggcgttact cgacagagga 5820
123 gagcaagaaa tggagccagt agatcctaga ctagagccct ggaagcatcc aggaagtcag 5880
124 cctaaaactg cttgtaccaa ttgctattgt aaaaagtgtt gctttcattg ccaagtttgt 5940
125 ttcatgacaa aagccttagg catctcctat ggcaggaaga agcggagaca gcgacgaaga 6000
126 gctcatcaga acagtcagac tcatcaagct tctctatcaa agcagtaagt agtacatgta 6060
127 atgcaaccta aatagtagc aatagtagca ttagtagtag caataataat agcaatagtt 6120
128 gtgtggtcca tagtaatcat agaatatagg aaaatattaa gacaaagaaa aatagacagg 6180
129 ttaattgata gactaataga aagagcagaa gacagtggca atgagagtga aggagaagta 6240
130 tcagcacttg tggagatggg ggtggaaatg gggcaccatg ctccttggga tattgatgat 6300
131 ctgtagtgct acagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaagga 6360
132 agcaaccacc actctatttt gtgcatcaga tgctaaagca tatgatacag aggtacataa 6420
133 tgtttgggcc acacatgcct gtgtacccac agaccccaac ccacaagaag tagtattggt 6480
134 aaatgtgaca gaaaatttta acatgtggaa aaatgacatg gtagaacaga tgcatgagga 6540
135 tataatcagt ttatgggatc aaagcctaaa gccatgtgta aaattaaccc cactctgtgt 6600
136 tagtttaaag tgcactgatt tgaagaatga tactaatacc aatagtagta gcgggagaat 6660
137 gataatggag aaaggagaga taaaaaactg ctctttcaat atcagcacaa gcataagaga 6720
138 taaggtgcag aaagaatatg cattctttta taaacttgat atagtaccaa tagataatac 6780
139 cagctatagg ttgataagtt gtaacacctc agtcattaca caggcctgtc caaaggtatc 6840
140 ctttgagcca attcccatac attattgtgc cccggctggt tttgcgattc taaaatgtaa 6900
141 taataagacg ttcaatggaa caggaccatg tacaaatgtc agcacagtac aatgtacaca 6960
142 tggaatcagg ccagtagtat caactcaact gctgttaaat ggcagtctag cagaagaaga 7020
143 tgtagtaatt agatctgcca atttcacaga caatgctaaa accataatag tacagctgaa 7080
144 cacatctgta gaaattaatt gtacaagacc caacaacaat acaagaaaaa gtatccgtat 7140
145 ccagagggga ccagggagag catttgttac aataggaaaa ataggaaata tgagacaagc 7200
146 acattgtaac attagtagag caaaatggaa tgccacttta aaacagatag ctagcaaatt 7260
147 aagagaacaa tttggaaata ataaaacaat aatctttaag caatcctcag gaggggaccc 7320
148 agaaattgta acgcacagtt ttaattgtgg aggggaattt ttctactgta attcaacaca 7380
149 actgtttaat agtacttggt ttaatagtac ttggagtact gaagggtcaa ataacactga 7440
150 aggaagtgac acaatcacac tcccatgcag aataaaacaa tttataaaca tgtggcagga 7500
151 agtaggaaaa gcaatgtatg cccctcccat cagtggacaa attagatgtt catcaaatat 7560
152 tactgggctg ctattaacaa gagatggtgg taataacaac aatgggtccg agatcttcag 7620
```

```
153 acctggagga ggcgatatga gggacaattg gagaagtgaa ttatataaat ataaagtagt 7680
154 aaaaattgaa ccattaggag tagcacccac caaggcaaag agaagagtgg tgcagagaga 7740
155 aaaaagagca gtgggaatag gagctttgtt ccttgggttc ttgggagcag caggaagcac 7800
156 tatgggctgc acgtcaatga cgctgacggt acaggccaga caattattgt ctgatatagt 7860
157 gcagcagcag aacaatttgc tgagggctat tgaggcgcaa cagcatctgt tgcaactcac 7920
158 agtctggggc atcaaacagc tccaggcaag aatcctggct gtggaaagat acctaaagga 7980
159 tcaacagctc ctggggattt ggggttgctc tggaaaactc atttgcacca ctgctgtgcc 8040
160 ttggaatgct agttggagta ataaatctct ggaacagatt tggaataaca tgacctggat 8100
161 ggagtgggac agagaaatta acaattacac aagcttaata cactccttaa ttgaagaatc 8160
162 gcaaaaccag caagaaaaga atgaacaaga attattggaa ttagataaat gggcaagttt 8220
163 gtggaattgg tttaacataa caaattggct gtggtatata aaattattca taatgatagt 8280
164 aggaggcttg gtaggtttaa gaatagtttt tgctgtactt tctatagtga atagagttag 8340
165 gcagggatat tcaccattat cgtttcagac ccacctccca atcccgaggg gacccgacag 8400
166 gcccgaagga atagaagaag aaggtggaga gagagacaga gacagatcca ttcgattagt 8460
167 gaacggatcc ttagcactta tctgggacga tctgcggagc ctgtgcctct tcagctacca 8520
168 ccgcttgaga gacttactct tgattgtaac gaggattgtg gaacttctgg gacgcagggg 8580
169 gtgggaagcc ctcaaatatt ggtggaatct cctacagtat tggagtcagg aactaaagaa 8640
170 tagtgctgtt aacttgctca atgccacagc catagcagta gctgagggga cagatagggt 8700
171 tatagaagta ttacaagcag cttatagagc tattcgccac atacctagaa gaataagaca 8760
172 gggcttggaa aggatttttgc tataagatgg gtggcaagtg gtcaaaaagt agtgtgattg 8820
173 gatggcctgc tgtaagggaa agaatgagac gagctgagcc agcagcagat ggggtgggag 8880
174 cagtatctcg agacctagaa aaacatggag caatcacaag tagcaataca gcagctaaca 8940
175 atgctgcttg tgcctggcta gaagcacaag aggaggaaga ggtgggtttt ccagtcacac 9000
176 ctcaggtacc tttaagacca atgacttaca aggcagctgt agatcttagc cacttttttaa 9060
177 aagaaaaggg gggactggaa gggctaattc actcccaaag aagacaagat atccttgatc 9120
178 tgtggatcta ccacacacaa ggctacttcc ctgattggca gaactacaca ccagggccag 9180
179 gggtcagata tccactgacc tttggatggt gctacaagct agtaccagtt gagccagata 9240
180 aggtagaaga ggccaataaa ggagagaaca ccagcttgtt acaccctgtg agcctgcatg 9300
181 gaatggatga ccctgagaga gaagtgttag agtggaggtt tgacagccgc ctagcatttc 9360
182 atcacgtggc ccgagagctg catccggagt acttcaagaa ctgctgacat cgagcttgct 9420
183 acaagggact ttccgctggg gactttccag ggaggcgtgg cctgggcggg actggggagt 9480
184 ggcgagccct cagatgctgc atataagcag ctgcttttttg cctgtactgg gtctctctgg 9540
185 ttagaccaga tctgagcctg ggagctctct ggctaactag ggaacccact gcttaagcct 9600
186 caataaagct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg tgactctggt 9660
187 aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagca            9709
191 <210> SEQ ID NO: 2
192 <211> LENGTH: 854
193 <212> TYPE: PRT
194 <213> ORGANISM: Human immunodeficiency virus
196 <220> FEATURE:
197 <223> OTHER INFORMATION: Envelope Polyprotein
199 <400> SEQUENCE: 2
200 Met Arg Val Lys Glu Lys Tyr Gln His Leu Trp Arg Trp Gly Trp Lys
201  1               5                   10                  15
203 Trp Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Thr Glu
204              20                  25                  30
206 Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
207          35                  40                  45
209 Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
```

```
210      50                    55                    60
212 Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
213 65                    70                    75                    80
215 Pro Gln Glu Val Val Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
216                    85                    90                    95
218 Lys Asn Asp Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
219                    100                   105                   110
221 Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ser
222              115                   120                   125
224 Leu Lys Cys Thr Asp Leu Lys Asn Asp Thr Asn Thr Asn Ser Ser Ser
225       130                   135                   140
227 Gly Arg Met Ile Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn
228 145                   150                   155                   160
230 Ile Ser Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Phe Phe
231                   165                   170                   175
233 Tyr Lys Leu Asp Ile Val Pro Ile Asp Asn Thr Ser Tyr Arg Leu Ile
234              180                   185                   190
236 Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
237              195                   200                   205
239 Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu
240       210                   215                   220
242 Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val
243 225                   230                   235                   240
245 Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln
246              245                   250                   255
248 Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Asp Val Val Ile Arg Ser
249              260                   265                   270
251 Ala Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Thr
252              275                   280                   285
255 Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser
256       290                   295                   300
258 Ile Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys
259 305                   310                   315                   320
261 Ile Gly Asn Met Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp
262              325                   330                   335
264 Asn Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly
265              340                   345                   350
267 Asn Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu
268       355                   360                   365
270 Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn
271       370                   375                   380
273 Ser Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr
274 385                   390                   395                   400
276 Glu Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys
277              405                   410                   415
279 Arg Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met
280              420                   425                   430
282 Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr
283       435                   440                   445
```

RAW SEQUENCE LISTING ERROR SUMMARY          DATE: 04/29/2002
PATENT APPLICATION:   US/09/913,159A          TIME: 11:20:41

Input Set : **A:\seq listing.txt**
Output Set: **N:\CRF3\04292002\I913159A.raw**

Please Note:
Use of **n** and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; N Pos. 99

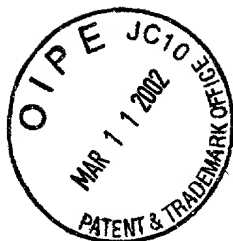**VERIFICATION SUMMARY**
PATENT APPLICATION:   **US/09/913,159A**

Input Set : **A:\seq listing.txt**
Output Set: **N:\CRF3\04292002\I913159A.raw**

L:10 M:270 C: Current Application Number differs, Replaced Application Number
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:399 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:60

<110> Strathmann AG & Co.

<120> Virus-Vaccine

<130> P057760

<140>
<141>

<150> 199 07 485.2
<151> 1999-02-12

<160> 12

<170> PatentIn Ver. 2.1

<210> 1
<211> 9709
<212> DNA
<213> Human immunodeficiency virus

<400> 1

```
tggaagggct aatttggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca  60
cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac  120
tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca  180
aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg  240
agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccgag  300
agctgcatcc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg  360
ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga gccctcagat  420
gctacatata agcagctgct ttttgcctgt actgggtctc tctggttaga ccagatctga  480
gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct  540
tgagtgctca agtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc  600
agacccttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacttgaaag  660
cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgctgaa gcgcgcacgg  720
caagaggcga ggggcggcga ctggtgagta cgccaaaaat tttgactagc ggaggctaga  780
aggagagaga tgggtgcgag agcgtcggta ttaagcgggg gagaattaga taaatgggaa  840
aaaattcggt taaggccagg gggaaagaaa caatataaac taaaacatat agtatgggca  900
agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaaggctgt  960
agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca  1020
ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc  1080
aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa  1140
gcagcagctg acacaggaaa caacagccag gtcagccaaa attaccctat agtgcagaac  1200
ctccaggggc aaatggtaca tcaggccata tcacctagaa cttttaaatgc atgggtaaaa  1260
gtagtagaag agaaggcttt cagcccagaa gtaatacccat gttttcagc attatcagaa  1320
ggagccaccc cacaagattt aaataccatg ctaaacacag tggggggaca tcaagcagcc  1380
atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca  1440
gtgcatgcag gcctattgc accaggccag atgagagaac caaggggaag tgacatagca  1500
ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca  1560
gtaggagaaa tctataaaag atggataatc ctgggattaa ataaaatagt aagaatgtat  1620
agccctacca gcattctgga cataagacaa ggaccaaagg aacccttag agactatgta  1680
gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg  1740
acagaaacct tgttggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg  1800
ggaccaggag cgacactaga agaaatgatg acagcatgtc agggagtggg gggacccggc  1860
cataaagcaa gagttttggc tgaagcaatg agccaagtaa caaatccagc taccataatg  1920
```

```
atacagaaag gcaattttag gaaccaaaga aagactgtta agtgtttcaa ttgtggcaaa 1980
gaagggcaca tagccaaaaa ttgcagggcc cctaggaaaa agggctgttg gaaatgtgga 2040
aaggaaggac accaaatgaa agattgtact gagagacagg ctaatttttt agggaagatc 2100
tggccttccc acaagggaag gccagggaat tttcttcaga gcagaccaga gccaacagcc 2160
ccaccagaag agagcttcag gtttggggaa gagacaacaa ctccctctca gaagcaggag 2220
ccgatagaca aggaactgta tcctttagct tccctcagat cactctttgg cagcgacccc 2280
tcgtcacaat aaagatagggg gggcaattaa aggaagctct attagataca ggagcagatg 2340
atacagtatt agaagaaatg aatttgccag gaagatggaa accaaaaatg atagggggaa 2400
ttggaggttt tatcaaagta ggacagtatg atcagatact catagaaatc tgcggacata 2460
aagctatagg tacagtatta gtaggaccta cacctgtcaa cataattgga agaaatctgt 2520
tgactcagat tggctgcact ttaaattttc ccattagtcc tattgagact gtaccagtaa 2580
aattaaagcc aggaatggat ggcccaaaag ttaaacaatg gccattgaca gaagaaaaaa 2640
taaaagcatt agtagaaatt tgtacagaaa tggaaaagga aggaaaaatt tcaaaaattg 2700
ggcctgaaaa tccatacaat actccagtat ttgccataaa gaaaaaagac agtactaaat 2760
ggagaaaatt agtagatttc agagaactta ataagagaac tcaagatttc tgggaagttc 2820
aattaggaat accacatcct gcagggttaa aacagaaaaa atcagtaaca gtactggatg 2880
tgggcgatgc atatttttca gttcccttag ataaagactt caggaagtat actgcattta 2940
ccatacctag tataaacaat gagacaccag ggattagata tcagtacaat gtgcttccac 3000
agggatggaa aggatcacca gcaatattcc agtgtagcat gacaaaaatc ttagagcctt 3060
ttagaaaaca aaatccagac atagtcatct atcaatacat ggatgatttg tatgtaggat 3120
ctgacttaga aatagggcag catagaacaa aaatagagga actgagacaa catctgttga 3180
ggtggggatt taccacacca gacaaaaaac atcagaaaga acctccattc ctttggatgg 3240
gttatgaact ccatcctgat aaatggacag tacagcctat agtgctgcca gaaaaggaca 3300
gctggactgt caatgacata cagaaattag tgggaaaatt gaattgggca agtcagattt 3360
atgcagggat taaagtaagg caattatgta aacttcttag gggaaccaaa gcactaacag 3420
aagtagtacc actaacagaa gaagcagagc tagaactggc agaaaacagg gagattctaa 3480
aagaaccggt acatggagtg tattatgacc catcaaaaga cttaatagca gaaatacaga 3540
agcaggggca aggccaatgg acatatcaaa tttatcaaga gccatttaaa aatctgaaaa 3600
caggaaaata tgcaagaatg aagggtgccc acactaatga tgtgaaacaa ttaacagagg 3660
cagtacaaaa aatagccaca gaaagcatag taatatgggg aaagactcct aaatttaaat 3720
tacccataca aaaggaaaca tgggaagcat ggtggacaga gtattggcaa gccacctgga 3780
ttcctgagtg ggagtttgtc aatacccctc ccttagtgaa gttatggtac cagttagaga 3840
aagaacccat aataggagca gaaactttct atgtagatgg ggcagccaat agggaaacta 3900
aattaggaaa agcaggatat gtaactgaca gaggaagaca aaaagttgtc cccctaacgg 3960
acacaacaaa tcagaagact gagttacaag caattcatct agctttgcag gattcgggat 4020
tagaagtaaa catagtgaca gactcacaat atgcattggg aatcattcaa gcacaaccag 4080
ataagagtga tcagagttta gtcagtcaaa taatagagca gttaataaaa aaggaaaaag 4140
tctacctggc atgggtacca gcacacaaag gaattggagg aaatgaacaa gtagatgggt 4200
tggtcagtgc tggaatcagg aaagtactat ttttagatgg aatagataag gcccaagaag 4260
aacatgagaa atatcacagt aattggagag caatggctag tgattttaac ctaccacctg 4320
tagtagcaaa agaaatagta gccagctgtg ataaatgtca gctaaaaggg gaagccatgc 4380
atggacaagt agactgtagc ccaggaatat ggcagctaga ttgtacacat ttagaaggaa 4440
aagttatctt ggtagcagtt catgtagcca gtggatatat agaagcagaa gtaattccag 4500
cagagacagg gcaagaaaca gcatacttcc tcttaaaatt agcaggaaga tggccagtaa 4560
aaacagtaca tacagacaat ggcagcaatt tcaccagtac tacagttaag gccgcctgtt 4620
ggtgggcggg gatcaagcag gaatttggca ttccctacaa tccccaaagt caaggagtaa 4680
tagaatctat gaataaagaa ttaaagaaaa ttataggaca ggtaagagat caggctgaac 4740
atcttaagac agcagtacaa atggcagtat tcatccacaa ttttaaaaga aaggggggga 4800
ttgggggggta cagtgcaggg gaaagaatag tagacataat agcaacagac atacaaacta 4860
aagaattaca aaaacaaatt acaaaaattc aaaattttcg ggtttattac agggacagca 4920
gagatccagt ttggaaagga ccagcaaagc tcctctggaa aggtgaaggg gcagtagtaa 4980
tacaagataa tagtgacata aaagtagtgc caagaagaaa agcaaagatc atcaggcatt 5040
atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattaacaca 5100
tggaaaagat tagtaaaaca ccatatgtat atttcaagga aagctaagga ctggttttat 5160
agacatcact atgaaagtac taatccaaaa ataagttcag aagtacacat cccactaggg 5220
gatgctaaat tagtaataac aacatattgg ggtctgcata caggagaaag agactggcat 5280
ttgggtcagg gagtctccat agaatggagg aaaaagagat atagcacaca agtagaccct 5340
```

```
gacctagcag accaactaat tcatctgcac tattttgatt gtttttcaga atctgctata 5400
agaaatacca tattaggacg tatagttagt cctaggtgtg aatatcaagc aggacataac 5460
aaggtaggat ctctacagta cttggcacta gcagcattaa taaaaccaaa acagataaag 5520
ccacctttgc ctagtgttag gaaactgaca gaggacagat ggaacaagcc ccagaagacc 5580
aagggccaca gagggagcca tacaatgaat ggacactaga gcttttagag gaacttaaga 5640
gtgaagctgt tagacatttt cctaggatat ggctccataa cttaggacaa catatctatg 5700
aaacttacgg ggatacttgg gcaggagtgg aagccataat aagaattctg caacaactgc 5760
tgtttatcca tttcagaatt gggtgtcgac atagcagaat aggcgttact cgacagagga 5820
gagcaagaaa tggagccagt agatcctaga ctagagccct ggaagcatcc aggaagtcag 5880
cctaaaactg cttgtaccaa ttgctattgt aaaaagtgtt gctttcattg ccaagtttgt 5940
ttcatgacaa aagccttagg catctcctat ggcaggaaga agcggagaca gcgacgaaga 6000
gctcatcaga acagtcagac tcatcaagct tctctatcaa agcagtaagt agtacatgta 6060
atgcaaccta taatagtagc aatagtagca ttagtagtag caataataat agcaatagtt 6120
gtgtggtcca tagtaatcat agaatatagg aaaatattaa gacaagaaa aatagacagg 6180
ttaattgata gactaataga aagagcagaa gacagtggca atgagagtga aggagaagta 6240
tcagcacttg tggagatggg ggtggaaatg gggcaccatg ctccttggga tattgatgat 6300
ctgtagtgct acagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaagga 6360
agcaaccacc actctatttt gtgcatcaga tgctaaagca tatgatacag aggtacataa 6420
tgtttgggcc acacatgcct gtgtacccac agaccccaac ccacaagaag tagtattggt 6480
aaatgtgaca gaaaatttta acatgtggaa aaatgacatg gtagaacaga tgcatgagga 6540
tataatcagt ttatgggatc aaagcctaaa gccatgtgta aaattaaccc cactctgtgt 6600
tagtttaaag tgcactgatt tgaagaatga tactaatacc aatagtagta gcgggagaat 6660
gataatggag aaaggagaga taaaaaactg ctctttcaat atcagcacaa gcataagaga 6720
taaggtgcag aaagaatatg cattctttta taaacttgat atagtaccaa tagataatac 6780
cagctatagg ttgataagtt gtaacacctc agtcattaca caggcctgtc caaaggtatc 6840
ctttgagcca attcccatac attattgtgc cccggctggt tttgcgattc taaaatgtaa 6900
taataagacg ttcaatggaa caggaccatg tacaaatgtc agcacagtac aatgtacaca 6960
tggaatcagg ccagtagtat caactcaact gctgttaaat ggcagtctag cagaagaaga 7020
tgtagtaatt agatctgcca atttcacaga caatgctaaa accataatag tacagctgaa 7080
cacatctgta gaaattaatt gtacaagacc caacaacaat acaagaaaaa gtatccgtat 7140
ccagagggga ccagggagag catttgttac aataggaaaa ataggaaata tgagacaagc 7200
acattgtaac attagtagag caaaatggaa tgccacttta aaacagatag ctagcaaatt 7260
aagagaacaa tttggaaata ataaaacaat aatctttaag caatcctcag gaggggaccc 7320
agaaattgta acgcacagtt ttaattgtgg aggggaattt ttctactgta attcaacaca 7380
actgtttaat agtacttggt ttaatagtac ttggagtact gaagggtcaa ataacactga 7440
aggaagtgac acaatcacac tcccatgcag aataaaacaa tttataaaca tgtggcagga 7500
agtaggaaaa gcaatgtatg cccctcccat cagtggacaa attagatgtt catcaaatat 7560
tactgggctg ctattaacaa gagatggtgg taataacaac aatgggtccg agatcttcag 7620
acctggagga ggcgatatga gggacaattg gagaagtgaa ttatataaat ataaagtagt 7680
aaaaattgaa ccattaggag tagcacccac caaggcaaag agaagagtgg tgcagagaga 7740
aaaaagagca gtgggaatag gagctttgtt ccttgggttc ttgggagcag caggaagcac 7800
tatgggctgc acgtcaatga cgctgacggt acaggccaga caattattgt ctgatatagt 7860
gcagcagcag aacaatttgc tgagggctat tgaggcgcaa cagcatctgt tgcaactcac 7920
agtctggggc atcaaacagc tccaggcaag aatcctggct gtggaaagat acctaaagga 7980
tcaacagctc ctggggattt ggggttgctc tggaaaactc atttgcacca ctgctgtgcc 8040
ttggaatgct agttggagta taaatctct ggaacagatt tggaataaca tgacctggat 8100
ggagtgggac agagaaatta acaattacac aagcttaata cactccttaa ttgaagaatc 8160
gcaaaaccag caagaaaaga tgaacaaga attattggaa ttagataaat gggcaagttt 8220
gtggaattgg tttaacataa caaattggct gtggtatata aaattattca taatgatagt 8280
aggaggcttg gtaggtttaa gaatagtttt tgctgtactt tctatagtga atagagttag 8340
gcagggatat tcaccattat cgtttcagac ccacctccca atcccgaggg gacccgacag 8400
gcccgaagga atagaagaag aaggtggaga gagagacaga gacagatcca ttcgattagt 8460
gaacggatcc ttagcactta tctgggacga tctgcggagc ctgtgcctct tcagctacca 8520
ccgcttgaga gacttactct tgattgtaac gaggattgtg gaacttctgg gacgcagggg 8580
gtgggaagcc ctcaaatatt ggtggaatct cctacagtat tggagtcagg aactaaagaa 8640
tagtgctgtt aacttgctca atgccacagc catagcagta gctgagggga cagataggt 8700
tatagaagta ttacaagcag cttatagagc tattcgccac atacctagaa gaataagaca 8760
```

```
gggcttggaa aggattttgc tataagatgg gtggcaagtg gtcaaaaagt agtgtgattg 8820
gatggcctgc tgtaagggaa agaatgagac gagctgagcc agcagcagat ggggtgggag 8880
cagtatctcg agacctagaa aaacatggag caatcacaag tagcaataca gcagctaaca 8940
atgctgcttg tgcctggcta gaagcacaag aggaggaaga ggtgggtttt ccagtcacac 9000
ctcaggtacc tttaagacca atgacttaca aggcagctgt agatcttagc cacttttaa  9060
aagaaaaggg gggactggaa gggctaattc actcccaaag aagacaagat atccttgatc 9120
tgtggatcta ccacacacaa ggctacttcc ctgattggca gaactacaca ccagggccag 9180
gggtcagata tccactgacc tttggatggt gctacaagct agtaccagtt gagccagata 9240
aggtagaaga ggccaataaa ggagagaaca ccagcttgtt acaccctgtg agcctgcatg 9300
gaatggatga ccctgagaga gaagtgttag agtggaggtt tgacagccgc ctagcatttc 9360
atcacgtggc ccgagagctg catccggagt acttcaagaa ctgctgacat cgagcttgct 9420
acaagggact ttccgctggg gactttccag ggaggcgtgg cctgggcggg actggggagt 9480
ggcgagccct cagatgctgc atataagcag ctgctttttg cctgtactgg gtctctctgg 9540
ttagaccaga tctgagcctg ggagctctct ggctaactag ggaacccact gcttaagcct 9600
caataaagct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg tgactctggt 9660
aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagca            9709
```

<210> 2
<211> 854
<212> PRT
<213> Human immunodeficiency virus

<220>
<223> Envelope Polyprotein

<400> 2
```
Met Arg Val Lys Glu Lys Tyr Gln His Leu Trp Arg Trp Gly Trp Lys
 1               5                  10                  15

Trp Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Thr Glu
                20                  25                  30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
            35                  40                  45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
        50                  55                  60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 65                  70                  75                  80

Pro Gln Glu Val Val Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
                85                  90                  95

Lys Asn Asp Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
                100                 105                 110

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ser
            115                 120                 125

Leu Lys Cys Thr Asp Leu Lys Asn Asp Thr Asn Thr Asn Ser Ser Ser
        130                 135                 140

Gly Arg Met Ile Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn
145                 150                 155                 160
```

Ile Ser Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Phe Phe
                165                 170                 175

Tyr Lys Leu Asp Ile Val Pro Ile Asp Asn Thr Ser Tyr Arg Leu Ile
            180             185             190

Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
        195             200             205

Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu
    210             215             220

Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val
225             230             235             240

Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln
            245             250             255

Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Asp Val Val Ile Arg Ser
            260             265             270

Ala Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Thr
        275             280             285

Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser
    290             295             300

Ile Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys
305             310             315             320

Ile Gly Asn Met Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp
            325             330             335

Asn Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly
            340             345             350

Asn Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu
            355             360             365

Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn
    370             375             380

Ser Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr
385             390             395             400

Glu Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys
            405             410             415

Arg Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met
            420             425             430

Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr
    435             440             445

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu

```
            450                 455                 460

Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
465                 470                 475                 480

Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro
                    485                 490                 495

Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly
                500                 505                 510

Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
            515                 520                 525

Gly Cys Thr Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser
    530                 535                 540

Asp Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln
545                 550                 555                 560

Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
                565                 570                 575

Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly
            580                 585                 590

Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp
        595                 600                 605

Asn Ala Ser Trp Ser Asn Lys Ser Leu Glu Gln Ile Trp Asn Asn Met
    610                 615                 620

Thr Trp Met Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser Leu Ile
625                 630                 635                 640

His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln
                645                 650                 655

Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asn
            660                 665                 670

Ile Thr Asn Trp Leu Trp Tyr Ile Lys Leu Phe Ile Met Ile Val Gly
        675                 680                 685

Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn
    690                 695                 700

Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro
705                 710                 715                 720

Ile Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly
                725                 730                 735

Glu Arg Asp Arg Asp Arg Ser Ile Arg Leu Val Asn Gly Ser Leu Ala
            740                 745                 750
```

Leu Ile Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg
    755                     760             765

Leu Arg Asp Leu Leu Leu Ile Val Thr Arg Ile Val Glu Leu Leu Gly
    770                 775             780

Arg Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr
785                 790             795                     800

Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Asn Leu Leu Asn Ala Thr
                805                 810                 815

Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Leu Gln
            820             825             830

Ala Ala Tyr Arg Ala Ile Arg His Ile Pro Arg Arg Ile Arg Gln Gly
        835                 840             845

Leu Glu Arg Ile Leu Leu
    850


<210> 3
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      oligonucleotide for cloning

<400> 3
aagatgtagt aattagatct gccaatttca cagacaatgc taaaaccata atagtacagc 60
tgaacacatc gttagaaatt aattgtacaa gacccaacaa caataca             107


<210> 4
<211> 120
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      oligonucleotide for cloning

<220>
<221> misc_feature
<222> (97)..(99)
<223> Sequence at this position: (GA)(AT)(GATC), ie.
      base at position 97 can be G or A, base at
      position 98 can be A or T, and base at
      position 99 can be G, A, T or C.

<400> 4
ttttgctcta gaaatgttac aatgtgcttg tcttatgtct cctgttgcag cttctgttgc 60
atgaaatgct ctccctggtc cgatatggat actatgrwnt tttcttgtat tgttgttggg 120

```
<210> 5
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
     sequencing primer

<400> 5
ccatgtacaa atgtcag                                                   17


<210> 6
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
     sequencing primer

<400> 6
aaaactgtgc gttacaa                                                   17


<210> 7
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
     sequencing primer

<400> 7
gtaaaacgac ggccagt                                                   17



<210> 8
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
     sequencing primer

<400> 8
caggaaacag ctatgac                                                   17


<210> 9
<211> 2148
```

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic DNA

<220>
<221> misc_feature
<222> (3)..(9)
<223> BstEII cleavage site

<220>
<221> misc_feature
<222> (2143)..(2148)
<223> BamHI cleavage site

<400> 9
```
tgggtcaccg tctattatgg ggtgcctgtg tggaaggaag caaccaccac tctattttgt 60
gcatcagatg ctaaagcata tgatacagag gtacataatg tttgggccac acatgcctgt 120
gtacccacag accccaaccc acaagaagta gtattggtaa atgtgacaga aaattttaac 180
atgtggaaaa atgacatggt agaacagatg catgaggata taatcagttt atgggatcaa 240
agccttaagc catgtgtaaa attaacccca ctctgtgtta gtttaaagtg cactgatttg 300
aagaatgata ctaataccaa tagtagtagc gggagaatga taatggagaa aggagagata 360
aaaaactgca gcttcaatat cagcacaagc ataagagata aggtgcagaa agaatatgca 420
ttcttttata aacttgatat agtaccaata gataatacca gctataggtt gataagttgt 480
aacacctcag tgatcacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 540
tattgtgccc cggctggttt tgcgattcta aaatgtaata ataagacgtt caatggaaca 600
ggaccatgta caaatgtcag cacagtacaa tgtacacatg gaattcgacc agtagtatca 660
actcaactgc tgttaaatgg cagtctagca gaagaagatg tagtaattag atctgccaat 720
ttcacagaca atgctaaaac cataatagta cagctgaaca catctgtaga aattaattgt 780
acaagaccca acaacaatac aagaaaaagt atccgtatcc agagggggacc agggagagca 840
tttgttacaa taggaaaaat aggaaatatg agacaagcac attgtaacat ttctagagca 900
aaatggaatg ccactttaaa acagatagct agcaaattaa gagaacaatt tggaaataat 960
aaaacaataa tctttaagca gtcatccgga ggggacccag aaattgtaac gcacagtttt 1020
aattgtggag gggaattttt ctactgtaat tcaacacaac tgtttaatag tacttggtttt 1080
aatagtactt ggagtactga agggtcaaat aacactgaag gaagtgacac aatcacactc 1140
ccatgcagaa taaaacaatt tataaacatg tggcaggaag taggaaaagc aatgtatgcc 1200
cctcccatca gtggccaaat tagatgttca tcaaatatta ctgggctgct attaactcga 1260
gatggtggta ataacaacaa tgggtccgag attttcagac ctggaggagg cgatatgagg 1320
gataattgga gaagtgaatt atataaatat aaagtagtaa aaattgaacc attaggagta 1380
gcacccacca aggcaaagag acgcgtggtg cagagagaaa agcgcgcagt gggaatagga 1440
gctctgttcc ttgggttctt gggagcagca ggaagcacta tgggcgcagc gtcaatgacg 1500
ctgacggtac aggccagaca attattgtct gatatagtgc agcagcagaa caatttgctg 1560
agggcaattg aggcgcaaca gcatctgttg caactcacag tctggggcat caaacagctc 1620
caggcaagaa tcctggctgt ggaaagatac ctaaaggatc aacagctcct ggggatttgg 1680
ggttgctctg gaaaactcat ttgcaccact gctgtgcctt ggaatgctag ttggagtaat 1740
aaatctctgg aacagatttg gaataacatg acctggatgg agtgggacag agaaattaac 1800
aattacacaa gcttaataca ctccttaatt gaagaatcgc aaaaccagca agaaaagaat 1860
gaacaagaat tattggaatt agataaatgg gcaagtttgt ggaattggtt taacataaca 1920
aattggctgt ggtatataaa attattcata atgatagtag gaggcttggt aggtttaaga 1980
atagtttttg ctgtactttc tatagtgaat agagttaggc agggatattc accattatcg 2040
tttcagaccc acctcccaat cccgagggga cccgacaggc ccgaaggaat agaagaagaa 2100
ggtggagaga gagacagaga cagatccatt cgattagtga acggatcc 2148
```

<210> 10
<211> 6229

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic DNA

<220>
<221> sig_peptide
<222> (1293)..(1295)
<223> env ATG

<220>
<221> misc_feature
<222> (1377)..(1379)
<223> env AGT, gp120 start

<220>
<221> misc_feature
<222> (1397)..(1403)
<223> BstEII cleavage site

<220>
<221> misc_feature
<222> (3537)..(3542)
<223> BamHI cleavage site

<220>
<221> misc_feature
<222> (3855)..(3857)
<223> env TAA, stop

<400> 10
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 60
ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg 120
ccacgttcgc cggctttccc cgtcaagctc taaatcgggg gctccctta gggttccgat 180
ttagtgcttt acggcacctc gaccccaaaa aacttgatta gggtgatggt tcacgtagtg 240
ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata 300
gtggactctt gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt 360
tataagggat tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat 420
ttaacgcgaa ttttaacaaa atattaacgc ttacaatttc cattcgccat tcaggctgcg 480
caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg 540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg 600
taaaacgacg gccagtgagc gtctagttat taatagtaat caattacggg gtcattagtt 660
catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 720
ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca 780
atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca 840
gtacatcaag tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg 900
cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc 960
tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt 1020
ggatagcggt ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 1080
ttgtttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg 1140
acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc tcgtttagtg 1200
aaccgtcaga tcgcctggag acgccatcca cgctgttttg acctccatag aagacaccgg 1260
gacaattcga gctcggtacc gtcgacgcca ccatgagagt gaaggagaag tatcagcact 1320
tgtggagatg ggggtggaaa tggggcacca tgctccttgg gatattgatg atctgtagtg 1380
ctacagaaaa attgtgggtc accgtctatt atggggtacc tgtgtggaag gaagcaacca 1440
ccactctatt ttgtgcatca gatgctaaag catatgatac agaggtacat aatgtttggg 1500

```
ccacacatgc ctgtgtaccc acagacccca acccacaaga agtagtattg gtaaatgtga 1560
cagaaaattt taacatgtgg aaaaatgaca tggtagaaca gatgcatgag gatataatca 1620
gtttatggga tcaaagccta aagccatgtg taaaattaac cccactctgt gttagtttaa 1680
agtgcactga tttgaagaat gatactaata ccaatagtag tagcgggaga atgataatgg 1740
agaaaggaga gataaaaaac tgctctttca atatcagcac aagcataaga gataaggtgc 1800
agaaagaata tgcattcttt tataaacttg atatagtacc aatagataat accagctata 1860
ggttgataag ttgtaacacc tcagtcatta cacaggcctg tccaaaggta tcctttgagc 1920
caattcccat acattattgt gccccggctg gttttgcgat tctaaaatgt aataataaga 1980
cgttcaatgg aacaggacca tgtacaaatg tcagcacagt acaatgtaca catggaatca 2040
ggccagtagt atcaactcaa ctgctgttaa atggcagtct agcagaagaa gatgtagtaa 2100
ttagatctgc caatttcaca gacaatgcta aaaccataat agtacagctg aacacatctg 2160
tagaaattaa ttgtacaaga cccaacaaca atacaagaaa aagtatccgt atccagaggg 2220
gaccagggag agcatttgtt acaataggaa aaataggaaa tatgagacaa gcacattgta 2280
acattagtag agcaaaatgg aatgccactt taaaacagat agctagcaaa ttaagagaac 2340
aatttggaaa taataaaaca ataatcttta agcaatcctc aggaggggac ccagaaattg 2400
taacgcacag ttttaattgt ggaggggaat ttttctactg taattcaaca caactgttta 2460
atagtacttg gtttaatagt acttggagta ctgaagggtc aaataacact gaaggaagtg 2520
acacaatcac actcccatgc agaataaaac aatttataaa catgtggcag gaagtaggaa 2580
aagcaatgta tgcccctccc atcagtggac aaattagatg ttcatcaaat attactgggc 2640
tgctattaac aagagatggt ggtaataaca caatgggtc cgagatcttc agacctggag 2700
gaggcgatat gagggacaat ggagaagtg aattatataa atataaagta gtaaaaattg 2760
aaccattagg agtagcaccc accaaggcaa agagaagagt ggtgcagaga gaaaaaagag 2820
cagtgggaat aggagctttg ttccttgggt tcttgggagc agcaggaagc actatgggct 2880
gcacgtcaat gacgctgacg gtacaggcca gacaattatt gtctgatata gtgcagcagc 2940
agaacaattt gctgagggct attgaggcgc aacagcatct gttgcaactc acagtctggg 3000
gcatcaaaca gctccaggca agaatcctgg ctgtggaaag atacctaaag gatcaacagc 3060
tcctggggat ttggggttgc tctggaaaac tcatttgcac cactgctgtg ccttggaatg 3120
ctagttggag taataaatct ctggaacaga tttggaataa catgacctgg atggagtggg 3180
acagagaaat taacaattac acaagcttaa tacactcctt aattgaagaa tcgcaaaacc 3240
agcaagaaaa gaatgaacaa gaattattgg aattagataa atgggcaagt ttgtggaatt 3300
ggtttaacat aacaaattgg ctgtggtata taaaattatt cataatgata gtaggaggct 3360
tggtaggttt aagaatagtt tttgctgtac tttctatagt gaatagagtt aggcagggat 3420
attcaccatt atcgtttcag acccacctcc caatcccgag gggacccgac aggcccgaag 3480
gaatagaaga agaaggtgga gagagagaca gagacagatc cattcgatta gtgaacggat 3540
ccttagcact tatctgggac gatctgcgga gcctgtgcct cttcagctac caccgcttga 3600
gagacttact cttgattgta acgaggattg tggaacttct gggacgcagg gggtgggaag 3660
ccctcaaata ttggtggaat ctcctacagt attggagtca ggaactaaag aatagtgctg 3720
ttaacttgct caatgccaca gccatagcag tagctgaggg gacagatagg gttatagaag 3780
tattacaagc agcttataga gctattcgcc acatacctag aagaataaga cagggcttgg 3840
aaaggatttt gctataagat gggtggcaag tggtcaaaaa gtagtgtgat tggatggcct 3900
gctgtaaggg aaagaatgag acgagctgag ccagcagcag atgggtgggg agcagtatct 3960
cgagatctag actagaacta gcttcgatcc agacatgata agatacattg atgagtttgg 4020
acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat 4080
tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca 4140
ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt aaaacctcta 4200
caaatgtggt atggctgatt atgatcctgc ctcgcgcgtt tcggtgatga cggtgaaaac 4260
ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga tgccgggagc 4320
agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggcgc agccatgacc 4380
cagtcacgta gcgatagcgg agtgtatact ggcttaacta tgcggcatca gagcagattg 4440
tactgagagt gcaccatatg tcgggccgcg ttgctggcgt ttttccatag gctccgcccc 4500
cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 4560
taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 4620
ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc 4680
tcacgctgta ggtatctcag ttcggtgtag tcgttcgct ccaagctggg ctgtgtgcac 4740
gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 4800
ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg 4860
aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga 4920
```

```
aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 4980
agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gttttttttgt ttgcaagcag 5040
cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggtct 5100
gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg 5160
atcttcacct agatccttttt aaattaaaaa tgaagtttta aatcaatcta agtatatat 5220
gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 5280
tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 5340
gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct 5400
ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca 5460
actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 5520
ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg 5580
tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc 5640
cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag 5700
ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg 5760
ccatccgtaa gatgctttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag 5820
tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat 5880
agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg 5940
atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca 6000
gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca 6060
aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat 6120
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag 6180
aaaaataaac aaatagggggt tccgcgcaca tttccccgaa aagtgccac          6229


<210> 11
<211> 860
<212> DNA
<213> Human immunodeficiency virus

<220>
<221> misc_feature
<222> (1)..(860)
<223> PI-932 original sequence V1-V2-V3-loop

<400> 11
tgtgtaccca cagacccccaa cccacaaaag gtagtattgg aaaatgtgac agaaaattt 60
aacatgtgga aaaatgacat ggtagaacag atgcatgagg atataatcaa tttatgggat 120
caaagcctaa agccatgtgt aaaactaacc ccactctgtg ttactttaaa ttgcactgat 180
gctgatttaa attgcaataa tactgatttta aattgcacta aagctaattt ggggaaaaat 240
actcataaca atactattag tgggaaaata atagagaaag tagaaataaa aaactgctct 300
ttcaaggtca ccacaggcat aagggataag atgcaaaaag aatatgcact tttgaataaa 360
cttgatatag taccaataga taatgataag aataatacta actttatatt gataagttgt 420
aacacctcga ccattacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 480
ttttgtgccc cggctggttt tgcgattcta aagtgtaatg aaaagagtta cagtggaaaa 540
ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattaggcc agtagtgtca 600
actcaactgc tgttgaatgg cagtctagca gaaaaagaag tagtaattag atctgagaat 660
ttcacagaca atgctaaaac cataatagta cagctgaagg aatctgtaaa cattacttgt 720
ataagacccc acaacactgt aacagacagg atacatatag gccagggag atcatttcat 780
acaacaagaa aaataaaagg agatataaga caagcacatt gtagccttag agaaaagat 840
tggaataaca cttttacaaga                                          860


<210> 12
<211> 870
<212> DNA
<213> Artificial Sequence
```

<220>

<223> Description of the artificial sequence: PI-932
    gene cassette, comprising the cleavage sites for
    restriction enzymes BspT1, PstI, BclI, EcoRI,
    BglII, PvuII, XbaII, NheI

<400> 12

```
tgtgtaccca cagaccccaa cccacaaaag gtagtattgg aaaatgtgac agaaaatttt 60
aacatgtgga aaaatgacat ggtagaacag atgcatgagg atataatcaa tttatgggat 120
caaagcctta agccatgtgt aaaactaacc ccactctgtg ttactttaaa ttgcactgat 180
gctgatttaa attgcaataa tactgattta aattgcacta aagctaattt ggggaaaaat 240
actcataact gcagtattag tgggaaaata atagagaaag tagaaataaa aaactgctct 300
ttcaaggtca ccacaggcat aagggataag atgcaaaaag aatatgcact tttgaataaa 360
cttgatatag taccaataga taatgataag aataatacta actttatatt gataagttgt 420
aacacctcgg tgatcacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 480
ttttgtgccc cggctggttt tgcgattcta aagtgtaatg aaaagagtta cagtggaaaa 540
ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattcggcc agtagtgtca 600
actcaactgc tgttgaatgg cagtctagca gaaaagaag tagtaattag atctgagaat 660
ttcacagaca atgctaaaac cataatagta cagctgaagg aatctgtaaa cattacttgt 720
ataagacccc acaacactgt aacagacagg atacatatag ggccagggag atcatttcat 780
acaacaagaa aaataaaagg agatataaga caagcacatt gtagcctttc tagaaaagat 840
tggaataaca ctttacaaga gatagctagc                                  870
```

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

SCHREIBER                                                    Atty. Ref.: 35-213

Serial No.  Unknown                                          Group:

U.S. National Phase of PCT/EP99/09759

Filed:  August 10, 2001                                      Examiner:

For:   VIRAL VACCINE

          *    *    *    *    *    *    *    *    *    *    *

                                                    August 10, 2001

Assistant Commissioner for Patents
Washington, DC  20231

Sir:

## LETTER
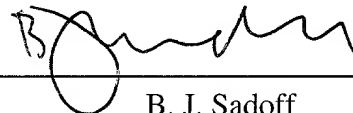
The attached paper and computer readable copies of the Sequence Listing are the same.

No new matter has been added.

                              Respectfully submitted,

                              **NIXON & VANDERHYE P.C.**


                    By: _____
                                       B. J. Sadoff
                                     Reg. No. 36,663


1100 North Glebe Road, 8th Floor
Arlington, VA 22201-4714
Telephone:  (703) 816-4000
Facsimile:  (703) 816-4100

550311

SEQUENCE LISTING

<110> Strathmann AG & Co.

<120> Virus-Vaccine

<130> P057760

<140>
<141>

<150> 199 07 485.2
<151> 1999-02-12

<160> 12

<170> PatentIn Ver. 2.1

<210> 1
<211> 9709
<212> DNA
<213> Human immunodeficiency virus

<400> 1

```
tggaagggct aatttggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca 60
cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac 120
tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca 180
aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg 240
agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccgag 300
agctgcatcc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg 360
ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga ccctcagat 420
gctacatata agcagctgct ttttgcctgt actgggtctc tctggttaga ccagatctga 480
gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct 540
tgagtgctca aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600
agaccctttt agtcagtgtg aaaatctct agcagtggcg cccgaacagg gacttgaaag 660
cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgctgaa gcgcgcacgg 720
caagaggcga ggggcggcga ctggtgagta cgccaaaaat tttgactagc ggaggctaga 780
aggagagaga tgggtgcgag agcgtcggta ttaagcgggg gagaattaga taatgggaa 840
aaaattcggt taaggccagg gggaaagaaa caatataaac taaaacatat agtatgggca 900
agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaaggctgt 960
agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca 1020
ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc 1080
aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa 1140
gcagcagctg acacaggaaa caacagccag gtcagccaaa attaccctat agtgcagaac 1200
ctccaggggc aaatggtaca tcaggccata tcacctagaa ctttaaatgc atgggtaaaa 1260
gtagtagaag agaaggcttt cagcccagaa gtaatacccca tgttttcagc attatcagaa 1320
ggagccaccc cacaagattt aaataccatg ctaaacacag tggggggaca tcaagcagcc 1380
atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca 1440
gtgcatgcag ggcctattgc accaggccag atgagagaac caaggggaag tgacatagca 1500
ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca 1560
gtaggagaaa tctataaaag atggataatc ctgggattaa ataaaatagt aagaatgtat 1620
agccctacca gcattctgga cataagacaa ggaccaaagg aacccttta g agactatgta 1680
gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg 1740
acagaaacct tgttggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg 1800
ggaccaggag cgacactaga agaaatgatg acagcatgtc agggagtggg gggacccggc 1860
cataaagcaa gagtttttggc tgaagcaatg agccaagtaa caaatccagc taccataatg 1920
atacagaaag gcaattttta g gaaccaaaga aagactgtta agtgtttcaa ttgtggcaaa 1980
gaagggcaca tagccaaaaa ttgcagggcc cctaggaaaa agggctgttg aaatgtgga 2040
aaggaaggac accaaatgaa agattgtact gagagacagg ctaatttttt aggg aagatc 2100
tggccttccc acaagggaag gccagggaat ttttcttcaga gcagaccaga gccaacagcc 2160
ccaccagaag agagcttcag gtttggggaa gagacaacaa ctccctctca gaagcaggag 2220
ccgatagaca aggaactgta tcctttagct tccctcagat cactctttgg cagcgacccc 2280
tcgtcacaat aaagataggg gggcaattaa aggaagctct attagataca ggagcagatg 2340
atacagtatt agaagaaatg aatttgccag gaagatggaa accaaaaatg ataggggggaa 2400
ttggaggttt tatcaaagta ggacagtatg atcagatact catagaaatc tgcggacata 2460
aagctatagg tacagtatta gtaggaccta cacctgtcaa cataattgga agaaatctgt 2520
tgactcagat tggctgcact ttaaattttc ccattagtcc tattgagact gtaccagtaa 2580
aattaaagcc aggaatggat ggcccaaaag ttaaacaatg gccattgaca gaagaaaaaa 2640
```

```
taaaagcatt agtagaaatt tgtacagaaa tggaaaagga aggaaaaatt tcaaaaattg 2700
ggcctgaaaa tccatacaat actccagtat ttgccataaa gaaaaaagac agtactaaat 2760
ggagaaaatt agtagatttc agagaactta ataagagaac tcaagatttc tgggaagttc 2820
aattaggaat accacatcct gcagggttaa aacagaaaa atcagtaaca gtactggatg 2880
tgggcgatgc atattttttca gttcccttag ataaagactt caggaagtat actgcattta 2940
ccatacctag tataaacaat gagacaccag ggattagata tcagtacaat gtgcttccac 3000
agggatggaa aggatcacca gcaatattcc agtgtagcat gacaaaaatc ttagagcctt 3060
ttagaaaaca aaatccagac atagtcatct atcaatacat ggatgatttg tatgtaggat 3120
ctgacttaga aatagggcag catagaacaa aaatagagga actgagacaa catctgttga 3180
ggtggggatt taccacacca gacaaaaaac atcagaaaga acctccattc ctttggatgg 3240
gttatgaact ccatcctgat aaatggacag tacagcctat agtgctgcca gaaaaggaca 3300
gctggactgt caatgacata cagaaattag tgggaaaatt gaattgggca agtcagattt 3360
atgcagggat taaagtaagg caattatgta aacttcttag gggaaccaaa gcactaacag 3420
aagtagtacc actaacagaa gaagcagagc tagaactggc agaaaacagg gagattctaa 3480
aagaaccggt acatggagtg tattatgacc catcaaaaga cttaataga gaaatacaga 3540
agcaggggca aggccaatgg acatatcaaa tttatcaaga gccatttaaa aatctgaaaa 3600
caggaaaata tgcaagaatg aagggtgccc acactaatga tgtgaaacaa ttaacagagg 3660
cagtacaaaa aatagccaca gaaagcatag taatatgggg aaagactcct aaatttaaat 3720
tacccataca aaaggaaaca tgggaagcat ggtggacaga gtattggcaa gccacctgga 3780
ttcctgagtg ggagtttgtc aatacccctc ccttagtgaa gttatggtac cagttagaga 3840
aagaacccat aataggagca gaaactttct atgtagatgg ggcagccaat agggaaacta 3900
aattaggaaa agcaggatat gtaactgaca gaggaagaca aaaagttgtc cccctaacgg 3960
acacaacaaa tcagaagact gagttacaag caattcatct agctttgcag gattcgggat 4020
tagaagtaaa catagtgaca gactcacaat atgcattggg aatcattcaa gcacaaccag 4080
ataagagtga atcagagtta gtcagtcaaa taatagagca gttaataaaa aaggaaaaag 4140
tctacctggc atgggtacca gcacacaaag gaattggagg aaatgaacaa gtagatgggt 4200
tggtcagtgc tggaatcagg aaagtactat ttttagatgg aatagataag gcccaagaag 4260
aacatgagaa atatcacagt aattggagag caatggctag tgattttaac ctaccacctg 4320
tagtagcaaa agaaaatagta gccagctgtg ataaatgtca gctaaaaggg gaagccatgc 4380
atggacaagt agactgtagc ccaggaatat ggcagctaga ttgtacacat ttagaaggaa 4440
aagttatctt ggtagcagtt catgtagcca gtggatatat agaagcagaa gtaattccag 4500
cagagacagg gcaagaaaca gcatacttcc tcttaaaatt agcaggaaga tggccagtaa 4560
aaacagtaca tacagacaat ggcagcaatt tcaccagtac tacagttaag gccgcctgtt 4620
ggtgggcggg gatcaagcag gaatttggca ttccctacaa tccccaaagt caaggagtaa 4680
tagaatctat gaataaagaa ttaaagaaaa ttataggaca ggtaagagat caggctgaac 4740
atcttaagac agcagtacaa atggcagtat tcatccacaa tttttaaaaga aaaggggggga 4800
ttggggggta cagtgcaggg gaaagaatag tagacataat agcaacagac atacaaacta 4860
aagaattaca aaaacaaatt acaaaaattc aaaattttcg ggtttattac agggacagca 4920
gagatccagt ttggaaagga ccagcaaagc tcctctggaa aggtgaaggg gcagtagtaa 4980
tacaagataa tagtgacata aaagtagtgc caagaagaaa agcaaagatc atcagggatt 5040
atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattaacaca 5100
tggaaaagat tagtaaaaca ccatatgtat atttcaagga aagctaagga ctggttttat 5160
agacatcact atgaaagtac taatccaaaa ataagttcag aagtacacat cccactaggg 5220
gatgctaaat tagtaataac aacatattgg ggtctgcata caggagaaag agactggcat 5280
ttgggtcagg gagtctccat agaatggagg aaaaagagat atagcacaca agtagaccct 5340
gacctagcag accaactaat tcatctgcac tattttgatt gtttttcaga atctgctata 5400
agaaatacca tattaggacg tatagttagt cctaggtgtg aatatcaagc aggacataac 5460
aaggtaggat ctctacagta cttggcacta gcagcattaa taaaaccaaa acagataaag 5520
ccacctttgc ctagtgttag gaaactgaca gaggacagat ggaacaagcc ccagaagacc 5580
aagggccaca gagggagcca tacaatgaat ggacactaga gcttttagag gaacttaaga 5640
gtgaagctgt tagacatttt cctaggatat ggctccataa cttaggacaa catatctatg 5700
aaacttacgg ggatacttgg gcaggagtgg aagccataat aagaattctg caacaactgc 5760
tgtttatcca tttcagaatt gggtgtcgac atagcagaat aggcgttact cgacagagga 5820
gagcaagaaa tggagccagt agatcctaga ctagagccct ggaagcatcc aggaagtcag 5880
cctaaaactg cttgtaccaa ttgctattgt aaaaagtgtt gctttcattg ccaagtttgt 5940
ttcatgacaa aagccttagg catctcctat ggcaggaaga agcggagaca gcgacgaaga 6000
gctcatcaga acagtcagac tcatcaagct tctctatcaa agcagtaagt agtacatgta 6060
atgcaaccta taatagtagc aatagtagca ttagtagtag caataatagc agcaatagtt 6120
gtgtggtcca tagtaatcat agaatatagg aaaatattaa gacaaagaaa aatagacagg 6180
ttaattgata gactaataga aagagcagaa gacagtggca atgagagtga aggagaagta 6240
tcagcacttg tggagatggg ggtggaaatg gggcaccatg ctccttggga tattgatgat 6300
ctgtagtgct acagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaagga 6360
agcaaccacc actctatttt gtgcatcaga tgctaaagca tatgatacag aggtacataa 6420
tgtttgggcc acacatgcct gtgtacccac agaccccaac ccacaagaag tagtattggt 6480
aaatgtgaca gaaaatttta acatgtggaa aaatgacatg gtagaacaga tgcatgagga 6540
tataatcagt ttatgggatc aaagcctaaa gccatgtgta aaattaaccc cactctgtgt 6600
tagtttaaag tgcactgatt tgaagaatga tactaatacc aatagtagta gcgggagaat 6660
```

```
gataatggag  aaaggagaga  taaaaaactg  ctctttcaat  atcagcacaa  gcataagaga  6720
taaggtgcag  aaagaatatg  cattctttta  taaacttgat  atagtaccaa  tagataatac  6780
cagctatagg  ttgataagtt  gtaacacctc  agtcattaca  caggcctgtc  caaaggtatc  6840
cttttgagcca  attcccatac  attattgtgc  cccggctggt  tttgcgattc  taaaatgtaa  6900
taataagacg  ttcaatggaa  caggaccatg  tacaaatgtc  agcacagtac  aatgtacaca  6960
tggaatcagg  ccagtagtat  caactcaact  gctgttaaat  ggcagtctag  cagaagaaga  7020
tgtagtaatt  agatctgcca  atttcacaga  caatgctaaa  accataatag  tacagctgaa  7080
cacatctgta  gaaattaatt  gtacaagacc  caacaacaat  acaagaaaaa  gtatccgtat  7140
ccagagggga  ccagggagag  catttgttac  aataggaaaa  ataggaaata  tgagacaagc  7200
acattgtaac  attagtagag  caaaatggaa  tgccacttta  aaacagatag  ctagcaaatt  7260
aagagaacaa  tttggaaata  ataaaacaat  aatctttaag  caatcctcag  gaggggaccc  7320
agaaattgta  acgcacagtt  ttaattgtgg  aggggaattt  ttctactgta  attcaacaca  7380
actgtttaat  agtacttggt  ttaatagtac  ttggagtact  gaagggtcaa  ataacactga  7440
aggaagtgac  acaatcacac  tcccatgcag  aataaaacaa  tttataaaca  tgtggcagga  7500
agtaggaaaa  gcaatgtatg  cccctcccat  cagtggacaa  attagatgtt  catcaaatat  7560
tactgggctg  ctattaacaa  gagatggtgg  taataacaac  aatgggtccg  agatcttcag  7620
acctggagga  ggcgatatga  gggacaattg  gagaagtgaa  ttatataaat  ataaagtagt  7680
aaaaattgaa  ccattaggag  tagcacccac  caaggcaaag  agaagagtgg  tgcagagaga  7740
aaaaagagca  gtgggaatag  gagctttgtt  ccttgggttc  ttgggagcag  caggaagcac  7800
tatgggctgc  acgtcaatga  cgctgacggt  acaggccaga  caattattgt  ctgatatagt  7860
gcagcagcag  aacaatttgc  tgagggctat  tgaggcgcaa  cagcatctgt  tgcaactcac  7920
agtctggggc  atcaaacagc  tccaggcaag  aatcctggct  gtggaaagat  acctaaagga  7980
tcaacagctc  ctggggattt  ggggttgctc  tggaaaactc  atttgcacca  ctgctgtgcc  8040
ttggaatgct  agttggagta  ataaatctct  ggaacagatt  tggaataaca  tgacctggat  8100
ggagtgggac  agagaaatta  acaattacac  aagcttaata  cactccttaa  ttgaagaatc  8160
gcaaaaccag  caagaaaaga  atgaacaaga  attattggaa  ttagataaat  gggcaagttt  8220
gtggaattgg  tttaacataa  caaattggct  gtggtatata  aaattattca  taatgatagt  8280
aggaggcttg  gtaggtttaa  gaatagtttt  tgctgtactt  tctatagtga  atagagttag  8340
gcagggatat  tcaccattat  cgtttcagac  ccacctccca  atcccgaggg  gacccgacag  8400
gcccgaagga  atagaagaag  aaggtggaga  gagagacaga  gacagatcca  ttcgattagt  8460
gaacggatcc  ttagcactta  tctgggacga  tctgcggagc  ctgtgcctct  tcagctacca  8520
ccgcttgaga  gacttactct  tgattgtaac  gaggattgtg  gaacttctgg  gacgcagggg  8580
gtgggaagcc  ctcaaatatt  ggtggaatct  cctacagtat  tggagtcagg  aactaaagaa  8640
tagtgctgtt  aacttgctca  atgccacagc  catagcagta  gctgagggga  cagataggtt  8700
tatagaagta  ttacaagcag  cttatagagc  tattcgccac  atacctagaa  gaataagaca  8760
gggcttggaa  aggattttgc  tataagatgg  gtggcaagtg  gtcaaaaagt  agtgtgattg  8820
gatggcctgc  tgtaagggaa  agaatgagac  gagctgagcc  agcagcagat  ggggtgggag  8880
cagtatctcg  agacctagaa  aaacatggag  caatcacaag  tagcaataca  gcagctaaca  8940
atgctgcttg  tgcctggcta  gaagcacaag  aggaggaaga  ggtgggtttt  ccagtcacac  9000
ctcaggtacc  tttaagacca  atgacttaca  aggcagctgt  agatcttagc  cactttttaa  9060
aagaaaaggg  gggactggaa  gggctaattc  actcccaaag  aagacaagat  atccttgatc  9120
tgtggatcta  ccacacacaa  ggctacttcc  ctgattggca  gaactacaca  ccagggccag  9180
gggtcagata  tccactgacc  tttggatggt  gctacaagct  agtaccagtt  gagccagata  9240
aggtagaaga  ggccaataaa  ggagagaaca  ccagcttgtt  acaccctgtg  agcctgcatg  9300
gaatggatga  ccctgagaga  gaagtgttag  agtggaggtt  tgacagccgc  ctagcatttc  9360
atcacgtggc  ccgagagctg  catccggagt  acttcaagaa  ctgctgacat  cgagcttgct  9420
acaagggact  ttccgctggg  gactttccag  ggaggcgtgg  cctgggcggg  actggggagt  9480
ggcgagccct  cagatgctgc  atataagcag  ctgctttttg  cctgtactgg  gtctctctgg  9540
ttagaccaga  tctgagcctg  ggagctctct  ggctaactag  ggaacccact  gcttaagcct  9600
caataaagct  tgccttgagt  gcttcaagta  gtgtgtgccc  gtctgttgtg  tgactctggt  9660
aactagagat  ccctcagacc  cttttagtca  gtgtggaaaa  tctctagca              9709
```

<210> 2
<211> 854
<212> PRT
<213> Human immunodeficiency virus

<220>
<223> Envelope Polyprotein

<400> 2

```
Met Arg Val Lys Glu Lys Tyr Gln His Leu Trp Arg Trp Gly Trp Lys
1               5                   10                  15

Trp Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Thr Glu
            20                  25                  30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
        35                  40                  45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
    50                  55              .   60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65                  70                  75                  80

Pro Gln Glu Val Val Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
                85                  90                  95

Lys Asn Asp Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
            100                 105                 110

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ser
        115                 120                 125

Leu Lys Cys Thr Asp Leu Lys Asn Asp Thr Asn Thr Asn Ser Ser Ser
    130                 135                 140

Gly Arg Met Ile Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn
145                 150                 155                 160

Ile Ser Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Phe Phe
                165                 170                 175

Tyr Lys Leu Asp Ile Val Pro Ile Asp Asn Thr Ser Tyr Arg Leu Ile
            180             -   185                 190

Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
        195                 200                 205

Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu
    210                 215                 220

Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val
225                 230                 235                 240

Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln
                245                 250                 255

Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Asp Val Val Ile Arg Ser
            260                 265                 270

Ala Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Thr
            275                 280                 285
```

Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser
290                     295                 300

Ile Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys
305                     310                 315                 320

Ile Gly Asn Met Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp
325                     330                 335

Asn Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly
340                     345                 350

Asn Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu
355                     360                 365

Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn
370                     375                 380

Ser Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr
385                     390                 395                 400

Glu Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys
405                     410                 415

Arg Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met
420                     425                 430

Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr
435                     440                 445

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu
450                     455                 460

Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
465                     470                 475                 480

Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro
485                     490                 495

Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly
500                     505                 510

Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
515              520                 525

Gly Cys Thr Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser
530                     535                 540

Asp Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln
545                     550                 555                 560

Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
565                     570                 575

Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly
580                     585                 590

Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp
595                     600                 605

Asn Ala Ser Trp Ser Asn Lys Ser Leu Glu Gln Ile Trp Asn Asn Met
610                     615                 620

```
Thr Trp Met Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser Leu Ile
625             630         635             640
```

```
His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln
                645             650             655
```

```
Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asn
            660             665             670
```

```
Ile Thr Asn Trp Leu Trp Tyr Ile Lys Leu Phe Ile Met Ile Val Gly
        675             680             685
```

```
Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn
    690             695             700
```

```
Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro
705             710             715             720
```

```
Ile Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly
            725             730             735
```

```
Glu Arg Asp Arg Asp Arg Ser Ile Arg Leu Val Asn Gly Ser Leu Ala
        740             745             750
```

```
Leu Ile Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg
        755             760             765
```

```
Leu Arg Asp Leu Leu Leu Ile Val Thr Arg Ile Val Glu Leu Leu Gly
    770             775             780
```

```
Arg Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr
785             790             795             800
```

```
Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Asn Leu Leu Asn Ala Thr
            805             810             815
```

```
Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Leu Gln
        820             825             830
```

```
Ala Ala Tyr Arg Ala Ile Arg His Ile Pro Arg Arg Ile Arg Gln Gly
        835             840             845
```

```
Leu Glu Arg Ile Leu Leu
850
```

```
<210> 3
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      oligonucleotide for cloning

<400> 3
aagatgtagt aattagatct gccaatttca cagacaatgc taaaaccata atagtacagc 60
tgaacacatc gttagaaatt aattgtacaa gacccaacaa caataca             107
```

'<210> 4
<211> 120
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      oligonucleotide for cloning

<220>
<221> misc_feature
<222> (97)..(99)
<223> Sequence at this position: (GA)(AT)(GATC), ie.
      base at position 97 can be G or A, base at
      position 98 can be A or T, and base at
      position 99 can be G, A, T or C.

<400> 4
ttttgctcta gaaatgttac aatgtgcttg tcttatgtct cctgttgcag cttctgttgc 60
atgaaatgct ctccctggtc cgatatggat actatgrwnt tttcttgtat tgttgttggg 120


<210> 5
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      sequencing primer

<400> 5
ccatgtacaa atgtcag                                                  17


<210> 6
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      sequencing primer

<400> 6
aaaactgtgc gttacaa                                                  17


<210> 7
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      sequencing primer

<400> 7
gtaaaacgac ggccagt                                                  17

<210> 8
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence:
      sequencing primer

<400> 8
caggaaacag ctatgac                                              17


<210> 9
<211> 2148
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic DNA

<220>
<221> misc_feature
<222> (3)..(9)
<223> BstEII cleavage site

<220>
<221> misc_feature
<222> (2143)..(2148)
<223> BamHI cleavage site

<400> 9
tgggtcaccg tctattatgg ggtgcctgtg tggaaggaag caaccaccac tctattttgt 60
gcatcagatg ctaaagcata tgatacagag gtacataatg tttgggccac acatgcctgt 120
gtacccacag accccaaccc acaagaagta gtattggtaa atgtgacaga aaattttaac 180
atgtggaaaa atgacatggt agaacagatg catgaggata taatcagttt atgggatcaa 240
agccttaagc catgtgtaaa attaacccca ctctgtgtta gtttaaagtg cactgatttg 300
aagaatgata ctaataccaa tagtagtagc gggagaatga taatggagaa aggagagata 360
aaaaactgca gcttcaatat cagcacaagc ataagagata aggtgcagaa agaatatgca 420
ttctttata aacttgatat agtaccaata gataatacca gctataggtt gataagttgt 480
aacacctcag tgatcacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 540
tattgtgccc cggctggttt tgcgattcta aatgtaata ataagacgtt caatggaaca 600
ggaccatgta caaatgtcag cacagtacaa tgtacacatg gaattcgacc agtagtatca 660
actcaactgc tgttaaatgg cagtctagca gaagaagatg tagtaattag atctgccaat 720
ttcacagaca atgctaaaac cataatagta cagctgaaca catctgtaga aattaattgt 780
acaagaccca acaacaatac aagaaaaagt atccgtatcc agaggggacc agggagagca 840
tttgttacaa taggaaaaat aggaaatatg agacaagcac attgtaacat ttctagagca 900
aaatggaatg ccactttaaa acagatagct agcaaattaa gagaacaatt tggaaataat 960
aaaacaataa tctttaagca gtcatccgga ggggacccag aaattgtaac gcacagtttt 1020
aattgtggag gggaattttt ctactgtaat tcaacacaac tgtttaatag tacttggttt 1080
aatagtactt ggagtactga agggtcaaat aacactgaag gaagtgacac aatcacactc 1140
ccatgcagaa taaaacaatt tataaacatg tggcaggaag taggaaaagc aatgtatgcc 1200
cctcccatca gtggccaaat tagatgttca tcaaatatta ctgggctgct attaactcga 1260
gatggtggta taacaacaa tgggtccgag attttcagac ctggaggagg cgatatgagg 1320
gataattgga gaagtgaatt atataaatat aaagtagtaa aaattgaacc attaggagta 1380
gcacccacca aggcaaagag acgcgtggtg cagagagaaa agcgcgcagt gggaatagga 1440
gctctgttcc ttgggttctt gggagcagca ggaagcacta tgggcgcagc gtcaatgacg 1500
ctgacggtac aggccagaca attattgtct gatatagtgc agcagcagaa caatttgctg 1560
agggcaattg aggcgcaaca gcatctgttg caactcacag tctggggcat caaacagctc 1620
caggcaagaa tcctggctgt ggaaagatac ctaaaggatc aacagctcct ggggatttgg 1680
ggttgctctg gaaaactcat ttgcaccact gctgtgcctt ggaatgctag ttggagtaat 1740
aaatctctgg aacagatttg gaataacatg acctggatgg agtgggacag agaaattaac 1800
aattacacaa gcttaataca ctccttaatt gaagaatcgc aaaaccagca agaaaagaat 1860
gaacaagaat tattggaatt agataaatgg gcaagtttgt ggaattggtt taacataaca 1920
aattggctgt ggtatataaa attattcata atgatagtag gaggcttggt aggtttaaga 1980
atagtttttg ctgtactttc tatagtgaat agagttaggc agggatattc accattatcg 2040
tttcagaccc acctcccaat cccgagggga cccgacaggc ccgaaggaat agaagaagaa 2100

ggtggagaga gagacagaga cagatccatt cgattagtga acggatcc                    2148

<210> 10
<211> 6229
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: synthetic DNA

<220>
<221> sig_peptide
<222> (1293)..(1295)
<223> env ATG

<220>
<221> misc_feature
<222> (1377)..(1379)
<223> env AGT, gp120 start

<220>
<221> misc_feature
<222> (1397)..(1403)
<223> BstEII cleavage site

<220>
<221> misc_feature
<222> (3537)..(3542)
<223> BamHI cleavage site

<220>
<221> misc_feature
<222> (3855)..(3857)
<223> env TAA, stop

<400> 10
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 60
ccgctacact tgccagcgcc ctagcgcccg ctccttttcgc tttcttccct tcctttctcg 120
ccacgttcgc cggctttccc cgtcaagctc taaatcgggg gctccctttta gggttccgat 180
ttagtgcttt acggcacctc gaccccaaaa aacttgatta gggtgatggt tcacgtagtg 240
ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata 300
gtggactctt gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt 360
tataagggat tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat 420
ttaacgcgaa ttttaacaaa atattaacgc ttacaatttc cattcgccat tcaggctgcg 480
caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg 540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg 600
taaaacgacg gccagtgagc gtctagttat taatagtaat caattacggg gtcattagtt 660
catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 720
ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca 780
atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca 840
gtacatcaag tgtatcatat gccaagtacg cccctattg acgtcaatga cggtaaatgg 900
cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc 960
tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt 1020
ggatagcggt ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 1080
ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg 1140
acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc tcgtttagtg 1200
aaccgtcaga tcgcctggag acgccatcca cgctgttttg acctccatag aagacaccgg 1260
gacaattcga gctcggtacc gtcgacgcca ccatgagagt gaaggagaag tatcagcact 1320
tgtggagatg ggggtggaaa tggggcacca tgctccttgg gatattgatg atctgtagtg 1380
ctacagaaaa attgtgggtc accgtctatt atggggtacc tgtgtggaag gaagcaacca 1440
ccactctatt ttgtgcatca gatgctaaag catatgatac agaggtacat aatgtttggg 1500
ccacacatgc ctgtgtaccc acagacccca acccacaaga agtagtattg gtaaatgtga 1560
cagaaaattt taacatgtgg aaaaatgaca tggtagaaca gatgcatgag gatataatca 1620
gtttatggga tcaaagccta aagccatgtg taaaattaac cccactctgt gttagtttaa 1680
agtgcactga tttgaagaat gatactaata ccaatagtag tagcgggaga atgataatgg 1740
agaaaggaga gataaaaaac tgctctttca atatcagcac aagcataaga gataaggtgc 1800

```
agaaagaata tgcattcttt tataaacttg atatagtacc aatagataat accagctata 1860
ggttgataag ttgtaacacc tcagtcatta cacaggcctg tccaaaggta tcctttgagc 1920
caattcccat acattattgt gccccggctg gttttgcgat tctaaaatgt aataataaga 1980
cgttcaatgg aacaggacca tgtacaaatg tcagcacagt acaatgtaca catggaatca 2040
ggccagtagt atcaactcaa ctgctgttaa atggcagtct agcagaagaa gatgtagtaa 2100
ttagatctgc caatttcaca gacaatgcta aaaccataat agtacagctg aacacatctg 2160
tagaaattaa ttgtacaaga cccaacaaca atacaagaaa aagtatccgt atccagaggg 2220
gaccagggag agcatttgtt acaataggaa aaataggaaa tatgagacaa gcacattgta 2280
acattagtag agcaaaatgg aatgccactt taaaacagat agctagcaaa ttaagagaac 2340
aatttggaaa taataaaaca ataatcttta agcaatcctc aggaggggac ccagaaattg 2400
taacgcacag ttttaattgt ggaggggaat ttttctactg taattcaaca caactgtttta 2460
atagtacttg gtttaatagt acttggagta ctgaagggtc aaataacact gaaggaagtg 2520
acacaatcac actcccatgc agaataaaac aatttataaa catgtggcag gaagtaggaa 2580
aagcaatgta tgcccctccc atcagtggac aaattagatg ttcatcaaat attactgggc 2640
tgctattaac aagagatggt ggtaataaca acaatgggtc cgagatcttc agacctggag 2700
gaggcgatat gagggacaat tggagaagtg aattatataa atataaagta gtaaaaattg 2760
aaccattagg agtagcaccc accaaggcaa agagaagagt ggtgcagaga gaaaaaagag 2820
cagtgggaat aggagctttg ttccttgggt tcttgggagc agcaggaagc actatgggct 2880
gcacgtcaat gacgctgacg gtacaggcca gacaattatt gtctgatata gtgcagcagc 2940
agaacaattt gctgagggct attgaggcgc aacagcatct gttgcaactc acagtctggg 3000
gcatcaaaca gctccaggca agaatcctgg ctgtggaaag atacctaaag gatcaacagc 3060
tcctggggat ttggggttgc tctggaaaac tcatttgcac cactgctgtg ccttggaatg 3120
ctagttggag taataaatct ctggaacaga tttggaataa catgacctgg atggagtggg 3180
acagagaaat taacaattac acaagcttaa tacactcctt aattgaagaa tcgcaaaacc 3240
agcaagaaaa gaatgaacaa gaattattgg aattagataa atgggcaagt ttgtggaatt 3300
ggtttaacat aacaaattgg ctgtggtata taaaattatt cataatgata gtaggaggct 3360
tggtaggttt aagaatagtt tttgctgtac tttctatagt gaatagagtt aggcagggat 3420
attcaccatt atcgtttcag acccacctcc caatcccgag gggacccgac aggcccgaag 3480
gaatagaaga agaaggtgga gagagagaca gagacagatc cattcgatta gtgaacggat 3540
ccttagcact tatctgggac gatctgcgga gcctgtgcct cttcagctac caccgcttga 3600
gagacttact cttgattgta acgaggattg tggaacttct gggacgcagg gggtgggaag 3660
ccctcaaata ttggtggaat ctcctacagt attggagtca ggaactaaag aatagtgctg 3720
ttaacttgct caatgccaca gccatagcag tagctgaggg gacagatagg gttatagaag 3780
tattacaagc agcttataga gctattcgcc acatacctag aagaataaga cagggcttgg 3840
aaaggatttt gctataagat gggtggcaag tggtcaaaaa gtagtgtgat tggatggcct 3900
gctgtaaggg aaagaatgag acgagctgag ccagcagcag atggggtggg agcagtatct 3960
cgagatctag actagaacta gcttcgatcc agacatgata agatacattg atgagtttgg 4020
acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat 4080
tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca 4140
ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt aaaacctcta 4200
caaatgtggt atggctgatt atgatcctgc ctcgcgcgtt tcggtgatga cggtgaaaac 4260
ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga tgccgggagc 4320
agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggcgc agccatgacc 4380
cagtcacgta gcgatagcgg agtgtatact ggcttaacta tgcggcatca gagcagattg 4440
tactgagagt gcaccatatg tcgggccgcg ttgctggcgt ttttccatag gctccgcccc 4500
cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 4560
taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 4620
ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc 4680
tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac 4740
gaacccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 4800
ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg 4860
aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga 4920
aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 4980
agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag 5040
cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct 5100
gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg 5160
atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta agtatatat 5220
gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 5280
tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 5340
gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct 5400
ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca 5460
actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 5520
ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg 5580
tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc 5640
cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag 5700
ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg 5760
ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag 5820
```

```
tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat 5880
agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg 5940
atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca 6000
gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca 6060
aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat 6120
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag 6180
aaaaataaac aaatagggggt tccgcgcaca tttccccgaa aagtgccac        6229
```

<210> 11
<211> 860
<212> DNA
<213> Human immunodeficiency virus

<220>
<221> misc_feature
<222> (1)..(860)
<223> PI-932 original sequence V1-V2-V3-loop

<400> 11
```
tgtgtaccca cagaccccaa cccacaaaag gtagtattgg aaaatgtgac agaaaatttt 60
aacatgtgga aaatgacat ggtagaacag atgcatgagg atataatcaa tttatgggat 120
caaagcctaa agccatgtgt aaaactaacc ccactctgtg ttactttaaa ttgcactgat 180
gctgatttaa attgcaataa tactgattta aattgcacta aagctaattt ggggaaaaat 240
actcataaca atactattag tgggaaaata atagagaaag tagaaataaa aaactgctct 300
ttcaaggtca ccacaggcat aagggataag atgcaaaaag aatatgcact tttgaataaa 360
cttgatatag taccaataga taatgataag aataatacta actttatatt gataagttgt 420
aacacctcga ccattacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 480
ttttgtgccc cggctggttt tgcgattcta aagtgtaatg aaaagagtta cagtggaaaa 540
ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattaggcc agtagtgtca 600
actcaactgc tgttgaatgg cagtctagca gaaaaagaag tagtaattag atctgagaat 660
ttcacagaca atgctaaaac cataatagta cagctgaagg aatctgtaaa cattacttgt 720
ataagacccc acaacactgt aacagacagg atacatatag gccagggag atcatttcat 780
acaacaagaa aaataaaagg agatataaga caagcacatt gtagccttag gagaaaagat 840
tggaataaca ctttacaaga                                       860
```

<210> 12
<211> 870
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of the artificial sequence: PI-932
      gene cassette, comprising the cleavage sites for
      restriction enzymes BspTl, PstI, BclI, EcoRI,
      BglII, PvuII, XbaII, NheI

<400> 12
```
tgtgtaccca cagaccccaa cccacaaaag gtagtattgg aaaatgtgac agaaaatttt 60
aacatgtgga aaatgacat ggtagaacag atgcatgagg atataatcaa tttatgggat 120
caaagcctta agccatgtgt aaaactaacc ccactctgtg ttactttaaa ttgcactgat 180
gctgatttaa attgcaataa tactgattta aattgcacta aagctaattt ggggaaaaat 240
actcataact gcagtattag tgggaaaata atagagaaag tagaaataaa aaactgctct 300
ttcaaggtca ccacaggcat aagggataag atgcaaaaag aatatgcact tttgaataaa 360
cttgatatag taccaataga taatgataag aataatacta actttatatt gataagttgt 420
aacacctcgg tgatcacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 480
ttttgtgccc cggctggttt tgcgattcta aagtgtaatg aaaagagtta cagtggaaaa 540
ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattcggcc agtagtgtca 600
actcaactgc tgttgaatgg cagtctagca gaaaaagaag tagtaattag atctgagaat 660
ttcacagaca atgctaaaac cataatagta cagctgaagg aatctgtaaa cattacttgt 720
ataagacccc acaacactgt aacagacagg atacatatag gccagggag atcatttcat 780
acaacaagaa aaataaaagg agatataaga caagcacatt gtagcctttc tagaaaagat 840
tggaataaca ctttacaaga gatagctagc                            870
```

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

SCHREIBER                                    Atty. Ref.: 35-213

Serial No.  Unknown                          Group:

U.S. National Phase of PCT/EP99/09759

Filed: August 10, 2001                       Examiner:

For:  VIRAL VACCINE

         \*   \*   \*   \*   \*   \*   \*   \*   \*   \*   \*

August 10, 2001

Assistant Commissioner for Patents
Washington, DC  20231

Sir:

### PRELIMINARY AMENDMENT

In order to place the above-identified application in better condition for

examination, please amend the application as follows:

### IN THE CLAIMS

Please substitute the following amended claims for corresponding claims

previously presented.  A copy of the amended claims showing current revisions is

attached.

3.  (Amended) Protein vaccine according to claim 1 , characterized in that it
comprises a mixture of GP120 proteins of HIV which in each case differ from each other
in their amino acid sequence in the region of the V2 loop and/or of the V3 loop.

548990

6. (Amended) DNA vaccine according to claim 4 , characterized in that the mixture contains $\geq 10^3$ and preferably $\geq 10^4$ DNA molecules which differ from each other in their nucleic acid sequence.

7. (Amended) DNA vaccine according to claim 4, characterized in that it codes for a mixture of structurally different GP120 proteins of HIV, in which the vaccine contains a mixture of DNA molecules, the nucleic acid sequences of which differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop.

11. (Amended) Nucleic acid sequence according to claim 9 , characterized in that the sequence is modified by the introduction of silent mutations.

12. (Amended) Nucleic acid sequence according to claim 9, characterized in that it contains the sequence given in SEQ ID NO: 9.

17. (Amended) Nucleic acid sequence according to claim 15 or a single-stranded nucleic acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12 , characterized in that the fragment/the fragments contain(s) inosine, a nucleic acid exchange or a mutation at 9 to 20 positions.

18. (Amended) Double-stranded DNA which comprises hybrids of the single-stranded nucleic acid sequence according to claim 15 with a single-stranded nucleic

548990

acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12.

27. (Amended) Expression vector, characterized in that it contains an inserted nucleic acid sequence according to claim 9.

33. (Amended) Vector mixture according to claim 30, characterized in that the plasmids can be expressed in *E. coli* as host cell.

34. (Amended) Vector mixture according to claim 30, characterized in that the plasmids can be expressed in eukaryotic cells, preferably in Cos, CHO or BHK cells, as host cells.

40. (Amended) Process according to claim 38 , characterized in that the nucleic acid sequence coding for a viral protein is the sequence according SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same.

41. (Amended) Process for the preparation of the vector mixture according to claim 33, characterized in that plasmids, the nucleic acid sequences of which in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop in each case through random distribution of the bases at the

548990

varied nucleotide positions, are ligated into a vector which can be expressed in host cells.

43.  (Amended) Process for the preparation of the host cells composing transforming *E.coli.* with a vector mixture according to claim 30 .

44.  (Amended) Process for the preparation of a protein vaccine which comprises a mixture of viral protein molecules, characterized in that the molecules are sequence variants of a single viral protein or of part of same, the mixture containing $\geq 10^2$ sequence variants, which is obtainable by expression of a plasmid-DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations , said process comprising cultivating host cells according to claim 35 under conditions which allow the expression of the mixture of viral protein sequence variants.

45.  (Amended) Process for the preparation of a DNA vaccine which codes for a mixture of structurally different virus proteins, characterized in that the vaccine contains a mixture of sequence variants of a viral DNA molecule or of part of same, the mixture containing $\geq 10^2$ DNA molecules which differ from each other in their nucleic acid sequence, where the mixture, because of the variation of nucleotide positions, contains randomly distributed sequence combinations wherein said process is carried out according to claim 41 , wherein the plasmids are ligated into a vector which can be expressed in host cells of the organism to be vaccinated.

46.  (Amended) A method of preparing a vaccine comprising forming a mixture of structurally different viral proteins which are sequence variants of a viral protein or of part of same, for the prevention and/or therapy of a virus infection in humans.

47.  (Amended) A method of preparing a vaccine comprising forming a mixture according to claim 23 for the prevention and/or therapy of a HIV infection in humans.

548990

48. (Amended) A method of preparing a vaccine comprising forming a mixture of DNA molecules which code for sequence variants of a viral protein or of part of same, for the prevention and/or therapy of a virus infection in humans.

49. (Amended) A method of preparing a vaccine comprising forming a nucleic acid mixture according to claim 19 for the prevention and/or therapy of a virus infection in humans.

50. (Amended) A method of preparing a vaccine comprising forming the nucleic acid mixture according to claim 19 for the expression in host cells selected from the group consisting of *E. coli*, Cos, CHO and BHK cells.

51. (Amended) A method of producing a protein mixture which comprises sequence variants of the GP120 protein, characterized in that it is a mixture of proteins which contain amino acid sequences which in each case differ from each other in the V2 loop and/or in the V3 loop, the mixture being obtainable by expression of a plasmid DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations; said method comprising expressing a vector mixture which contains a mixture of plasmids which contains an inserted double-stranded DNA which comprises hybrids of the single-stranded nucleic acid sequence according to claim 15 with a single-stranded nucleic acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12; characterized in that the nucleic acid sequences of the plasmids differ in each case from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop, where the mixture of the plasmids, because

of the variation of nucleotide positions, contains randomly distributed sequence combinations .

52. (Amended) A method of preparing a protein mixture which comprises sequence variants of the GP120 protein, characterized in that it is a mixture of proteins which contain amino acid sequences which in each case differ from each other in the V2 loop and/or in the V3 loop, the mixture being obtainable by expression of a plasmid DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations, said method comprising culturing a host cell according to claim 35 .

54. (Amended) Pharmaceutical composition for the prevention and/or therapy of a virus infection, characterized in that it comprises a protein mixture and a nucleic acid mixture, the protein mixture comprising sequence variants of a viral protein or of part of same, and the nucleic acid mixture comprising DNA molecules which code for sequence variants of a viral protein or of part of same, said composition comprising a protein mixture according to claim 23 and a nucleic acid mixture which comprises double-stranded DNAs, the nucleic acid sequences of which are derived from the *env* sequence in SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same, characterized in that the nucleic acid sequences in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop .

## REMARKS

The above amendments are made to place the claims in a more traditional format. Attached hereto is a marked-up version of the changes made to the claims by the current amendment. The attached pages are captioned "**Version With Markings To Show Changes Made**."

548990

**SCHREIBER**
Serial No. **Unknown**
U.S. National Phase of PCT/EP99/09759

Respectfully submitted,

**NIXON & VANDERHYE P.C.**

By: _____

B. J. Sadoff
Reg. No. 36,663

BJS:ms
1100 North Glebe Road, 8th Floor
Arlington, VA 22201-4714
Telephone: (703) 816-4000
Facsimile: (703) 816-4100

548990

## VERSION WITH MARKINGS TO SHOW CHANGES MADE

## IN THE CLAIMS

3. (Amended) Protein vaccine according to claim 1 [or 2], characterized in that it comprises a mixture of GP120 proteins of HIV which in each case differ from each other in their amino acid sequence in the region of the V2 loop and/or of the V3 loop.

6. (Amended) DNA vaccine according to claim 4 [or 5], characterized in that the mixture contains $\geq 10^3$ and preferably $\geq 10^4$ DNA molecules which differ from each other in their nucleic acid sequence.

7. (Amended) DNA vaccine according to [one of claims 4 to 6] claim 4, characterized in that it codes for a mixture of structurally different GP120 proteins of HIV, in which the vaccine contains a mixture of DNA molecules, the nucleic acid sequences of which differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop.

11. (Amended) Nucleic acid sequence according to claim 9 [or 10], characterized in that the sequence is modified by the introduction of silent mutations.

12. (Amended) Nucleic acid sequence according to [claims 9 to 11]claim 9, characterized in that it contains the sequence given in SEQ ID NO: 9.

17. (Amended) Nucleic acid sequence according to claim 15 or [16] a single-stranded nucleic acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-

548990

long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12, characterized in that the fragment/the fragments contain(s) inosine, a nucleic acid exchange or a mutation at 9 to 20 positions.

18. (Amended) Double-stranded DNA which comprises hybrids of the single-stranded nucleic acid sequence according to claim 15 [or 17] with [the] a single-stranded nucleic acid sequence [according to claim 16 or 17] which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-Nhel fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12.

27. (Amended) Expression vector, characterized in that it contains an inserted nucleic acid sequence according to [claims 9 to 14]claim 9.

33. (Amended) Vector mixture according to [one of claims 30 to 32]claim 30, characterized in that the plasmids can be expressed in *E. coli* as host cell.

34. (Amended) Vector mixture according to [one of claims 30 to 32]claim 30, characterized in that the plasmids can be expressed in eukaryotic cells, preferably in Cos, CHO or BHK cells, as host cells.

548990

40. (Amended) Process according to claim 38 [or 39], characterized in that the nucleic acid sequence coding for a viral protein is the sequence according SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same.

41. (Amended) Process for the preparation of the vector mixture according to [claims 33 and 34]claim 33, characterized in that plasmids, the nucleic acid sequences of which in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop in each case through random distribution of the bases at the varied nucleotide positions, are ligated into a vector which can be expressed in host cells.

43. (Amended) Process for the preparation of the host cells [according to claim 35 or 36, characterized in that the host cells are transformed] composing transforming *E.coli.* with a vector mixture according to [claims] claim 30 [to 32].

44. (Amended) Process for the preparation of a protein vaccine which comprises a mixture of viral protein molecules, characterized in that the molecules are sequence variants of a single viral protein or of part of same, the mixture containing $\geq 10^2$ sequence variants, which is obtainable by expression of a plasmid-DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations [according to one of claims 1 to 3, characterized in that the host cells are cultivated] , said process comprising cultivating host cells according to [one of claims 35 to 37] claim 35 under conditions which allow the expression of the mixture of viral protein sequence variants.

45. (Amended) Process for the preparation of a DNA vaccine which codes for a mixture of structurally different virus proteins, characterized in that the vaccine contains a mixture of sequence variants of a viral DNA molecule or of part of same, the mixture containing $\geq 10^2$ DNA molecules which differ from each other in their nucleic acid sequence, where the mixture, because of the variation of nucleotide positions, contains

randomly distributed sequence combinations wherein said [according to [one of claims 4 to 8, characterized in that the] process is carried out according to claim 41 [or 42], wherein the plasmids [according to the invention being] are ligated into a vector which can be expressed in host cells of the organism to be vaccinated.

46. (Amended) A method of preparing a vaccine comprising forming a [Use of a] mixture of structurally different viral proteins which are sequence variants of a viral protein or of part of same, for [the preparation of a vaccine for] the prevention and/or therapy of a virus infection in humans.

47. (Amended) A method of preparing a vaccine comprising forming a mixture [Use of a protein mixture] according to [one of claims 23 to 25]claim 23 [for the preparation of a vaccine] for the prevention and/or therapy of a HIV infection in humans.

48. (Amended) A method of preparing a vaccine comprising forming [Use of] a mixture of DNA molecules which code for sequence variants of a viral protein or of part of same, for [the preparation of a vaccine for] the prevention and/or therapy of a virus infection in humans.

49. (Amended) A method of preparing a vaccine comprising [Use of] forming a nucleic acid mixture according to [claims 19 to 22]claim 19 for the [preparation of a vaccine for the] prevention and/or therapy of a virus infection in humans.

50. (Amended) A method of preparing a vaccine comprising [Use of] forming the nucleic acid mixture according to [one of claims 19 to 22]claim 19 for the [preparation of a vector mixture according to one of claims 30 to 32 which can be expressed] expression in host cells[, the host cells being] selected from the group consisting of E. coli, Cos, CHO and BHK cells.

51. (Amended) A method of producing a protein mixture  which comprises sequence variants of the GP120 protein, characterized in that it is a mixture of proteins

which contain amino acid sequences which in each case differ from each other in the V2 loop and/or in the V3 loop, the mixture being obtainable by expression of a plasmid DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations; said method comprising [Use of the vector mixture according to one of claims 30 to 32] expressing a vector mixture which contains a mixture of plasmids which contains an inserted double-stranded DNA which comprises hybrids of the single-stranded nucleic acid sequence according to claim 15 with a single-stranded nucleic acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, characterized in that in the V3 loop a 247 bp-long Bg1II-XbaI fragment or a 283 bp-long Bg1II-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-Bc1I fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to a nucleic acid sequence characterized in that it contains the sequence given in SEQ ID NO: 12; characterized in that the nucleic acid sequences of the plasmids differ in each case from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop, where the mixture of the plasmids, because of the variation of nucleotide positions, contains randomly distributed sequence combinations [for the expression of a protein mixture according to one of claims 23 to 25].

52. (Amended) A method of preparing a protein mixture which comprises sequence variants of the GP120 protein, characterized in that it is a mixture of proteins which contain amino acid sequences which in each case differ from each other in the V2 loop and/or in the V3 loop, the mixture being obtainable by expression of a plasmid DNA mixture which, because of the variation of nucleotide positions, contains randomly distributed sequence combinations, said method comprising culturing a [Use of the] host cell according to [one of claims 35 to 37]claim 35 [for the preparation of a protein mixture according to one of claims 23 to 25].

54. (Amended) Pharmaceutical composition for the prevention and/or therapy of a virus infection, characterized in that it comprises a protein mixture and a nucleic acid mixture, the protein mixture comprising sequence variants of a viral protein or of part of same, and the nucleic acid mixture comprising DNA molecules which code for sequence variants of a viral protein or of part of same, said composition comprising [according to claim 53, characterized in that it comprises] a protein mixture according to claim 23 [claims 23 to 25] and a nucleic acid mixture which comprises double-stranded DNAs, the nucleic acid sequences of which are derived from the *env* sequence in SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same, characterized in that the nucleic acid sequences in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop [according to claims 19 to 22].

- 13 -

548990

Strathmann AG & Co.

Sellhopsweg 1


22459 Hamburg

## VIRUS VACCINE


The present invention relates to a pharmaceutical
composition or a vaccine which comprises a mixture of viral
protein molecules which are sequence variants of a single
viral protein or of part of same. The invention furthermore
relates inter alia to a DNA vaccine which codes for a
mixture of structurally different virus proteins, the
vaccine containing a mixture of sequence variants of a
viral DNA molecule or of part of same, which code for
sequence variants of a viral protein or part. According to
a preferred version of the invention, the viral proteins
are sequence variants of the GP120 protein of the human
immunodeficiency virus (HIV) which differ from each other
in their amino acid sequence in the region of the V2 loop
and/or of the V3 loop, preferably both the V2 and the V3
loop. The invention furthermore relates to the preparation
of the virus vaccines including the intermediate stages or
constructs, preparation processes and uses connected with
them.


In many viral infections, in particular in the HIV-1
infection, a strong immune defence is observed, which is in

a position to control the virus over a period of several years. The period in which the virus is controlled and no illness symptoms are observed, is called the asymptomatic phase of the (HIV) illness. During the course of the illness, new virus variants repeatedly form. It is thereby possible for the virus to escape the human immune defence and to repeatedly infect new defence cells of the immune system (cf. M. Schreiber et al., J. Virol. 68 No. 6 (1994) 3908-3916; J. Gen. Virol. 77 (1996) 2403-2414; Clin. Exp. Immunol. 107 (1997) 15-20; J. Virol. 71 No. 12 (1997) 9198-9205).

In the prior art, for the treatment of viral illnesses, such as e.g. the polio virus (Horaud F et al., Biologicals, 1993, 21:311-316), Hanta virus (Ulrich R et al., 1998 Vaccine 16:272-280; Schmaljohn CS et al., 1992 vaccine 10:10-13), Lassa virus (Morrison HG et al., 1989 Virology 171:179-188; Clegg JC et al., 1987 Lancet 2:186-188), hepatitis A virus (Clemens et al., 1995 J. Infect. Dis. 171:Suppl.1:p.44-p.49; Andre et al., 1990 Prog. Med. Virol. 37:72-95) and hepatitis B virus infection (McAleer et al., 1984 Nature 307:178-180) but also the HIV infection (Egan et al., 1995 J. Infect. Dis. 171:1623-1627, Kovac et al., 1993 J. Clin. Invest. 92:919-928), only single, genetically unmodified, specific virus antigens or single inactivated virus strains have been used for studies of suitable vaccination strategies. In the case of HIV, for example, the external envelope proteins of two virus strains were up until now prepared for vaccine experiments on humans (MN and SF2) (Zolla-Pazner et al., J. Infect. Dis. 178 (1998) 1502-1506). Both GP120 molecules differ from each other in the amino acid sequence, but particularly in the amino acid

sequence of the variable regions, such as e.g. of the V3 loop. Because of the different V3 loop sequences, the two virus strains have different phenotypic properties. The HIV-1 MN strain is a virus variant which preferentially infects macrophages and cells which possess the viral co-receptor CCR5. Viruses of the SF2 type, on the other hand, multiply preferentially in T cells and use the viral co-receptor CXCR4. Such viruses are therefore also called T cell trophs (e.g. HIV strain) or macrophagotrophs (e.g. HIV strain MN). Vaccination experiments have been carried out in chimpanzees with these two GP120 variants. It has been shown in these experiments that an immune response can be induced which not only protects against the two HIV strains MN and SF2, but is also capable of preventing infection by other virus strains, but also by HIV-1 patient isolates. In contrast to chimpanzees, only one of the two GP120 variants has been used in humans for vaccination experiments up until now. Neither MN GP120 nor SF2 GP120 offers certain protection against infection by a virus variant such as occurs in a person infected by HIV-1 (patient isolate or wild-type virus). J. A. Levy offers in *"HIV and the Pathogenesis of AIDS"*, Publisher: Jay. A. Levy, Chapter 15, 2nd Edition, ASM Press Washington, D.C., 1998, an overview of the state of research in connection with vaccination with GP120 envelope protein.

The disadvantage of the vaccination strategies discussed or examined up until now in the prior art is, among other things, that, with the vaccines used, one is not in a position to prevent the formation of ever-new virus variants during the course of the viral illness.

The object of the present invention is therefore to provide a vaccine which in particular is in a position to prevent or clearly reduce the formation of new virus variants during the course of a viral illness and to limit or prevent the possibilities that the virus can escape the human immune defence and repeatedly infect new defence cells of the immune system.

In order to achieve the object, the items described in the following claims are proposed according to the invention.

The subject of the present invention is thus inter alia a protein vaccine which comprises a mixture of viral protein molecules, the molecules being sequence variants of a single viral protein or of part of same.

According to the invention, sequence variants of a protein are understood to be those molecules which contain an amino acid sequence derived vis-à-vis a native viral protein or a part (fragment) of same, the variants differing from each other in that at least one amino acid can be exchanged at any chosen point of the sequence or parts of same. The sequence variants preferably contain several amino acid exchanges at various points of the sequence, which are responsible for the production but also the bonding of virus-neutralizing antibodies. The number and position of the exchanged amino acids depends on the amino acid variability of the regions of the gp120 which are observed during the cell culture of adapted and wild-type HIV isolates. According to the invention, the sequence variants have a heterogeneity at at least two amino acid positions of the sequence or parts of same. A heterogeneity at three

to eight and preferably at more than eight amino acid
positions is particularly preferred, all amino acids
occurring at these positions being possible. The number of
possible sequence variants which represent the vaccine in
their totality result from the combination of the various
amino acids which are possible per position.

The invention relates in particular to a vaccine which
comprises a mixture of $\geq 10^2$ sequence variants, i.e. a
mixture of more than $10^2$ molecules of various amino acid
sequences (homologues). A vaccine is particularly preferred
which comprises $\geq 10^3$ and preferably $\geq 10^4$ sequence
variants.

Within the framework of the present invention it is assumed
that the vaccine can comprise, along with the sequence
variants, the protein per se from which the sequence
variants are derived.

An anti-viral active ingredient is thus provided according
to the invention which comprises virus-specific proteins
which all differ in their sequences or parts of same. In
order to achieve this, the protein-coding gene is newly
synthesized in order to generate new cleavage sites for
DNA-restriction enzymes, in order to permit the exchange of
specific regions. By chemical synthesis of DNA fragments
which code for the concerned protein section, a gene bank
is then produced for the protein-coding sequence. The
expression vectors which code for the protein, are then
transfected as a mixture into cells. From these protein-
producing cells the mixture of the various viral proteins

can then be isolated which according to the invention represents the preferred active ingredient.

In the case of HIV as a viral illness, the chronic infection and the associated continuous damaging of the immune system during the course of the illness lead to the complete loss of the specific virus defence. Neutralizing antibodies which have the property of forming a specific, very strong bond with certain antigenic structures are part of the specific virus defence. Foreign antigens are thereby marked and their interaction with e.g. virus receptors are blocked. Such a specific virus defence are neutralizing antibodies against the V3 loop of the external GP120 envelope protein. Such anti-V3 loop antibodies are in a position to prevent the infection of cells. In animal models, it has been shown that an infection by HIV-1 can be prevented by administering a specific monoclonal V3 loop antibody. By administering the same monoclonal antibody, it was also possible to heal an existing infection.

In contrast to experimental systems, the development of ever-new virus variants is however observed in the natural course of HIV infection. Thus hundreds of different V3 loop variants are found in a single patient at one time, as the variation of the HIV-1 is particularly high precisely in the V3 loop. The V3 loop is an important dominant antigenic domain of the GP120. A very specific humoral immune response is therefore formed against each V3 loop. The result of the induction of a highly specific immune response against the V3 loop is that neutralizing antibodies against the V3 loop of the HIV-1 variant A are not in a position to neutralize the variant B and vice

versa (Schreiber et al., J. Virol. 68 (1994) 3908-16,
Abrahamsson et al., 4 (1990) 107-12).

In the asymptomatic phase of the HIV illness, the cell-free
virus variants in the serum are neutralized by antibodies.
Only later can virus variants be observed in the serum,
which escape autologous neutralization. These V3 loop
escape variants are no longer recognized by V3 loop
antibodies and are thus no longer neutralized either. All
other virus variants which are found in the serum of the
patients are however recognized by autologous V3 loop
antibodies. This indicates that during the course of the
illness, virus variants occur against which no V3 loop
antibodies exist. If the antibody content of serum samples
of a patient is examined over a period of several years,
the V3 loop antibodies against the V3 loop escape variant
which occurs in the later stage can be detected in the
serum samples which were taken at the beginning of the
asymptomatic phase. The result is not, as with other
infectious diseases, the standard escape of the pathogen
through ever-new antigenic variation, but the switching-off
of the neutralizing immune response to a virus replicating
in the patient. The observed absence of the neutralizing V3
loop antibodies is thus the result of a continuous loss of
the type-specific V3 loop antibodies. The corresponding
virus can multiply through the loss of the neutralizing
antibodies, which leads to an increase in the viral load in
the serum of the patients. In the course of the illness,
the increase in the viral load in the lymph nodes is also
observed (Chun et al., Nature 387 (1997) 183-188).

A virus which establishes itself against the immune system
in the patient and develops into the dominant virus variant
must necessarily overcome all antiviral barriers which the
immune system offers. Along with neutralizing antibodies,
cytotoxic T cells are also in a position to suppress the
virus reproduction. Cytotoxic T cells effect the death of
HIV infected CD4+ T cells. Along with the loss of the
neutralizing antibodies, the loss of such cytotoxic T cells
against HIV infected cells is also observed (Zerhoui et
al., Thymus 24 (1997) 203-219; Gould et al., Semin Cell Dev
Biol 9 (1998) 321-328; Wagner et al., J Gen Virol 74 (1993)
1261-1269; O'Toole et al., AIDS Res Hum Retrovirsues 8
(1992) 1361-1368; Shearer et al., 137 (1986) 2514-2521).
The object of a heterologous vaccine is therefore to induce
or activate the mixture of many different GP120 envelope
proteins of HIV both neutralizing antibodies and cytotoxic
T cells against many different virus variants.

As already stated, the HIV illness is characterized in that
the virus is constantly changing. This is effected by the
error rate of the viral enzyme reverse transcriptase which
commits errors during the increase of the viral genetic
information. Mutants are thereby produced which are on the
one hand selected because of their biological properties
and the antiviral action of the human immune system. As
with HIV, there are also other pathogens which transcribe
their genetic information without error correction, which
can lead to the formation of many variants. These include
Hanta and Lassa viruses as well as hepatitis A, B and C
viruses. Thus it is observed in people vaccinated for
hepatitis B that vaccination fails in approx. 2%. In these

2%, hepatitis B "escape" variants are found which cannot be suppressed by the immune response induced by the vaccine.

The formation of vaccine, therapy and immune escape variants is based on the genetic variability of these viruses (Blum, Int J Clin Lab Res 27 (1997) 213-214; Jongerius et al., Transfusion 38 (1998) 56-59).

Up until now it has not been known in the state of the art how such a loss of virus-neutralizing antibodies can be countered. The strategies adopted to date were successful only in certain respects, i.e. in relation to the formation and the maintenance of individual variant-specific antibodies against specific viruses, but the treatment methods proposed to date in the form of antiviral medicaments or inoculants based on a memo-substance cannot counter the loss of type-specific antibodies against the variety of the protein variants formed by the virus in the course of the illness.

It is possible, through the present invention of a heterogeneous vaccine, to prevent the loss of neutralizing V3 antibodies or at least to clearly counter this loss.

On the basis of the present invention, for the first time an immunoreconstitutive treatment of virus-infected patients, in particular infected with HIV-1, is provided in which the immune system is activated such that the naturally acquired immune defence against the virus population of the various virus variants is regenerated and stimulated anew, in order to thus prevent the virus variants present in the patient from multiplying.

In short, the invention is based on the concept of preparing a mixture of GP120 proteins for an active ingredient against HIV-1, these GP120 proteins differing e.g. either in the V2 or the V3 loop or simultaneously in the two variable domains, i.e. the V2 loop and the V3 loop. In order to achieve this, the GP120-coding gene has to be newly synthesized accompanied by the formation of new (monovalent) cleavage sites for DNA-restriction enzymes which allow the specific exchange of the e.g. V2 and V3 region. By chemical synthesis of DNA fragments which code for the V2 loop and V3 loop a V2/V3 gene bank for GP120 is then produced. The GP120 expressions vectors are finally transfected into cells as a mixture, and the mixture of the various GP120 proteins which can be used as an active ingredient for the prophylaxis and/or therapy of the HIV illness or AIDS can then be isolated from these GP120-producing cells.

The consideration on which the present invention is based thus comprises the preparation for an immunotherapy or a vaccine against HIV-1 of many different GP120 molecules which carry V3 loop sequences (and/or V2 loop sequences) such as can also be identified in virus variants in patients.

The production of a mixture of natural virus variants in the cell culture system is very expensive. It must be borne in mind in particular that the composition of a virus mixture changes as cells which can be infected by HIV-1 have to be used in the cell culture. Some virus variants which have selective advantages, e.g. faster growth

kinesis, therefore become dominant in the cell culture system after only a few days and expel the other slower virus variants. This applies above all when the various virus variants which are present in the peripheral mononuclear blood cells (PMBC) or in the serum of patients are to be isolated. As it is not possible to cultivate a constant mixture of HIV variants, neither can a corresponding mixture of GP120 proteins be produced nor isolated in this way from virus cultures.

Within the framework of the present invention, a genetic-engineering approach was therefore chosen for the preparation of HIV-1 variants and GP120 variants. For the preparation of virus variants and recombinant GP120, a cloning system was constructed which consists of two vectors. The central component of both vectors is a novel, chemically synthesized HIV-1 envelope protein gene (HIV-1 env gene) which codes for the GP120 protein. One vector contains the entire genome of HIV-1. After transfection of this vector into cells, the cells produce infectious virus particles. This facilitates the production of a mixture of virus variants as cells can be used which are resistant to an HIV-1 infection. Therefore, in such a cell culture system, the composition of the virus mixture cannot be changed, because of the biological properties of the virus variants. The second vector facilitates the expression of GP120-variants in eukaryotic cells. This vector serves directly for the preparation of the inoculant (the vaccine).

For the genetic-engineering preparation and expression of the V3 loop GP120 variants in eukaryotic cells, a special

gene construct for HIV-1-*env* was created, called gene cassette in the following. For the preparation of the gene cassette, according to the invention the entire coding sequence of the *env* gene was chemically synthesized. With the help of this method it is possible to change the coding sequence of the gene as desired, new DNA-recognition sequences for DNA-cutting restriction enzymes being able to be incorporated into the *env*-gene sequence. It must be borne in mind that the amino acid sequence of the original GP120 (preferably of the strains NL4-3 and PI-932) does not change, while new restriction cleavage sites form in the DNA sequence of the *env* gene. At the same time, according to the invention cleavage sites for enzymes which repeatedly occur in the *env* sequence are removed so that in each case there is only one cleavage site for a specific restriction enzyme, such as e.g. BglII. According to a preferred version of the invention ten new, once-occurring (monovalent) cleavage sites are inserted at intervals of approx. 150 base pairs. The new *env* gene produced in this way is also called gene cassette within the framework of the present invention. In principle, the new *env* fragment is similar to a polylinker.

The cleavage sites for the restriction enzymes BstEII and BamHI define the gene cassette preferred according to the invention (SEQ ID NO: 9). In order to be able to exchange the region of the *env* gene, both cleavage sites are present only once in the pBSCenvATG expression vector (SEQ ID NO: 10) for gp120 and in retroviral vector pNL4-3. The vector, given the reference code pBSCenv-V3 by the depositor, was deposited on 6th January 1999 at the DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH,

Mascheroder Weg 1b, 38124 Brunswick, Germany, under the accession number DSM 12612 in accordance with the Budapest Treaty.

Through the exchange of the gene section defined by BstEII and BamHI, the newly created sections of the *env* gene which are likewise defined by BstEII and BamHI, can be cloned into both vectors. The PstI/BclI V2 loop and the Bg1II/XbaI V3 loop *env* fragment are chemically or enzymatically synthesized and then cloned into a standard vector such as e.g. pUC 18 or 19. In this vector, the fragment is checked by sequencing and then cut out of the vector by a Pst/BcII or BgIII/XbaI digestion and transferred into the pBSCenvATG and then into the pNL4-3 vector. In place of the intermediate cloning into a standard vector (pUC 18, pUC 19 etc.), the PstI/BclI V2 loop and BglII/XbaI V3 loop fragments can also be cloned directly into pBSCenvATG.

As there is a cleavage site for a restriction enzyme every ca. 100 base pairs in the gene cassette, all regions of the *env* gene, in particular the V3 loop or V2 loop, can be exchanged for any DNA fragments. The section coding for the V3 loop can be exchanged for a Bg1II-XbaI fragment, which has a size of 244 base pairs (cf. SEQ ID NO: 9; nucleotides 708 to 955). The aim is to exchange the gene sections which code for the variable loops of the GP120 protein for new DNA fragments. The fragments are synthetically produced and then cloned into the *env* gene cassette. If the sequence is varied at certain positions during the synthesis of the V2 loop and V3 loop DNA fragments, i.e. all four nucleotides are incorporated equally or the nucleotide inosine is used at this position, *env*-gene variants with a pre-set

variation of the thereby-coded amino acid sequence can be produced. The V3 loop sequences of patient isolates are the starting point for the preparation of the GP120 mixture. If these variants are cloned into the *gp120* gene cassette, the corresponding GP120 proteins can be expressed with the variable antigenic domains in eukaryotic cells. Sequence data of HIV-1 patient isolates are available for the variation of the V3 loop region of the *env* gene. Patient isolates are virus variants which have been cultivated direct from patient material, serum or infected cells from the blood in the laboratory. These viruses are characterized, unlike HIV-1 viruses which are cultivated in the laboratory for a long time, by particular properties. These viruses are more resistant to neutralizing antibodies and chemokines. In contrast to cell culture-adapted viruses, patient isolates use different co-receptors for the infection of cells. As patient isolates differ greatly from laboratory-adapted viruses, the collected V2 loop or V3 loop sequence data of the *env* genes of such viruses are to be used for the preparation of the GP120 mixture. Within the framework of the invention, additional base exchanges are possible in regions outside the sequence sections coding for V2 and V3.

According to a particular version of the invention, a protein vaccine is made available which comprises a mixture of GP120 proteins of the human immunodeficiency virus (HIV) which in each case differ from each another in their amino acid sequence in the region of the V3 loop and/or of the V2 loop.

Within the framework of the present invention, during the
genetic-engineering preparation of the protein mixture or
the protein vaccine, sequence variations are introduced
preferably in regions of the V3 loops which lie outside
consensus sequences (cf. e.g. M. Schreiber et al., J.
Virol. 71 no. 12 (1997) 9198-9205). The consensus sequences
are sequence sections which essentially remain both between
different virus strains and during the formation of virus
variants as a viral illness progress (HIV illness) in the
body. Such a section is the sequence Gly-Pro-Gly-Arg-Ala-
Phe (GPGRAF) in the case of the HIV-1 of B sub-type. To the
left and right of this sequence unit lie short regions 4-10
amino acids long which can vary greatly. In Fig. 1 this is
represented using as example the V3 loop sequence
variations of patient isolates.

In addition to variations in the region of the V3 loop, the
molecules of the protein vaccine according to the invention
can also contain sequence variations in the region of the
V2 loop. Changes in the amino acid sequence outside
consensus sequence regions are preferred in this case also
(see Fig. 2). Furthermore, additional amino acid exchanges
are possible in regions outside the V2 and V3 loops.

In connection with the present invention the normal one- or
three-character codes are always used to describe the amino
acids. When there is variation of the amino acid sequences
inside the protein vaccine according to the invention, any
desired amino acid exchanges are possible. In each case,
however, amino acids of the viral GP120 sequence or of the
sequences of the V2 loop and V3 loop are preferably
exchanged for amino acids which also occur in other virus

variants at the corresponding sequence positions (see Fig. 2, position of 1997 sequence data in: Human Retroviruses and AIDS, A compilation and analysis of nucleic acid and amino acid sequences, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A., Editors: B. Korber, B. Hahn, B. Foley, J.W. Mellors, T. Leitner, G. Myers, F. McCutchan, C. Kuiken).

Within the framework of the present invention, for the first time a protein vaccine is made available which is prepared by genetic engineering and comprises sequence variants of specific viral antigens (proteins). The preparation of the vaccine is described in the following using the V3 loop as an example. It is self-evident to the person skilled in the art that the synthesis principle described below can be transferred to other viruses or viral proteins or parts of their sequences.

In short, a mixture of GP120 proteins with variable V3 sequences which are modelled on the variability of the V3 loop of the GP120 protein of HIV is prepared as a basis for the vaccine.

For this, the sequence coding for the GP120 protein is firstly cloned piecewise into the plasmid pUC 18/19. This takes place through a polylinker exchange so that all interesting sections with monovalent restriction sites are flanked by the possible restriction sites which were previously inserted in the gp120 sequence by silent mutations, and thus can be subsequently cut out and exchanged.

Two homologous nucleic acid oligomers are then prepared
with the help of a DNA synthesizer (length approx. 300 base
pairs). After hybridization has taken place, the double-
stranded DNA fragment forms with the DNA sequence of the V3
loop. Various standard methods from molecular biology can
be used for the preparation of the double-stranded V3 loop
DNA fragment. In one method, the variables are introduced
by inosines and in the case of the corresponding
complementary strand the variables are added by a
nucleotide mix (AGCT, AGC, AG, ...). In this method, the
DNA fragment is prepared by exclusively chemical means. In
another method, a combination of chemical and enzymatic DNA
synthesis, the variables are introduced by nucleotide
mixtures. The oligonucleotides are then hybridized at
complementary ends of a length of 20-30 bases, which are
not variable. The synthesis to produce the completely
double-stranded molecule takes place enzymatically using a
DNA polymerase. Both isothermal DNA polymerases (Klenow
fragment) or thermostable DNA polymerases (Taq-
polymerase)can be used. If Taq polymerase is used, then
larger amounts for the cloning of the V3 loop DNA fragment
can also be prepared using the polymerase chain reaction.
Using these methods, a double-stranded DNA fragment is
prepared which is variable at specific positions. This
degenerated DNA sequence codes for the corresponding
variety of V3 loop amino acid sequences of the GP120
protein mixture to be prepared, the protein vaccine.

The mixture of the synthesized V2 fragments has a PstI
cleavage site at the 5' end and a BclI cleavage site at the
3' end. The mixture of the synthesized V3 fragments has a
BglII cleavage site at the 5' end and an XbaI cleavage site

at the 3' end. Using the two respective cleavage sites, the fragment mixture is cloned into a vector, e.g. pUC18 delta-*env* or pUC 18 BstEII-BamHI (Fig. 3). After this cloning a mixture or a pool of plasmid DNAs of the vector pUC 18 forms, which all have the complete BstEII-BamHI *env* fragment. All plasmid DNAs differ exclusively in the sequence of the *env* V2 loop or V3 loop. By inserting the V2 or the V3 fragments, the deletion (delta *env*) of the *env* gene section in the vector pUC18 is stopped.

This thus-prepared mixture of the plasmid DNA with variable *env* fragments is transformed into *E. coli* and fermented. *E. coli* amplifies and replicates this mixture of plasmid DNA. All possible variables of these V2/V3 loop plasmids are produced.

This plasmid pool is then isolated. The *env* fragment is then cut out of the mixture of the plasmid DNA by BstEII and BamHI digestion and cloned direct into the vector for the expression of the viral *gp120*. This vector has the advantage that the GP120 is also expressed in the eukaryotic cell system.

This pool of BSCenvATG-vector DNA with variable *gp120* construct is now transfected into Cos or Chinese hamster ovary cells (CHO cells). These eukaryotes then express this pool of plasmids so that statistically the corresponding protein is translated to each variable and then – depending on which eukaryotes are used – correspondingly glycosylated. The proteins are then harvested, followed by purification to give the finished product (protein mixture or GP120 mixture with variable amino acid sequence).

By way of illustration, the process for producing the protein vaccine according to the invention is represented clearly in Fig. 4.

As already shown, variations in the V2 loop and/or in the V3 loop can be carried out. The variation of the V2 loop can take place according to the same scheme as described for V3.

The preparation of the protein vaccine according to the invention takes place, as described above, via various key constructs, i.e. nucleic acid intermediate steps and DNA constructs, which are essential for carrying out the invention.

As already explained, the nucleic acid sequence coding for the GP120 protein is cloned piecewise into pUC 18/19, the starting point according to the invention being a gp120 sequence in which the repeatedly present restriction sites are firstly modified by silent mutations such that each of the remaining restriction sites still occurs only once in the sequence. This sequence modification is necessary in order to be able to insert the expression cassette, which is necessary for producing the sequence variants, selectively at a quite specific point which is defined by two restriction sites each occurring only once in the sequence. Methods for the introduction of silent mutations in a nucleic acid sequence are well known to a person skilled in the art.

A subject of the present invention is therefore also a nucleic acid sequence which is derived from the *env* sequence in SEQ ID NO: 1 or a fragment of same, it being modified such that it contains exclusively monovalent restriction sites. Preferably, the modification takes place by the introduction of silent mutations. According to a preferred version of the invention, the nucleic acid sequence has the sequence shown in SEQ ID NO: 9. The sequence of the gene cassette can be modified such that it contains the total *env* gene or parts of the *env* gene of virus isolates from patients. Preferably, such a gene cassette is to contain the *env* sequence of the patient isolate PI-932 (SEQ ID NO: 11). A gene cassette preferred according to the invention based on the PI-932 sequence is shown in SEQ ID NO: 12.

A subject of the present invention is therefore also a single-stranded nucleic acid sequence which contains the region coding for the V3 loop and/or for the V2 loop or fragments or parts of same in which, in the case of the V3 loop, a Bg1II-XbaI (247 bp) or also a Bg1II-NheI (283 bp) fragment can be exchanged for a modified fragment which carries nucleic acid exchanges or mutations at least 6, preferably at 9 to 20 positions, and in the case of the V2 loop, a PstI-BclI (139 bp) or also a PstI-EcoRI (339 bp) fragment for a modified fragment which carries nucleic acid exchanges or mutations at least 6, preferably at 9 to 20 positions. The nucleotides within a nucleic acid sequence are in each case replaced either by inosine or by a mixture of 2-4 nucleotides. This is explained using the example in Fig. 5. If various amino acids are to be combined at 7 amino acid positions of the V3 loop 21 to give 1152

variants, two nucleotides must be introduced by chemical synthesis at each of 11 nucleic acid positions into the sequence of the single-stranded nucleic acids (oligonucleotides) (Fig. 5).

The single-stranded nucleic acid sequences are - as already mentioned above - hybridized or synthesized to give the double strand.

The present invention therefore furthermore relates to double-stranded DNA which comprises hybrids of the above-mentioned single-stranded nucleic acid sequences, where in each case one nucleic acid sequence (5'-3' oligomer or 3'-5' oligomer) contains inosines at one or more selected positions, and the other nucleic acid sequence (3'-5' oligomer or 5'-3' oligomer) in each case contains two, three or four of the possible nucleotides (adenine, A; thymine, T; guanine, G; cytosine, C) at the corresponding complementary positions during the later hybridization. Sequences are thereby produced in which the four bases are present randomly distributed at the respective positions of the single-stranded nucleic acid sequence (DNA) so that the desired combinations of A, T, G or C are produced for the corresponding desired amino acid codons (a codon is formed from a series of three nucleotides). Because of the variation of nucleotide positions, randomly distributed sequence combinations result. Therefore, e.g. for the combination of (A, C) (ACT) (ACGT) a total of 2x3x4=24 possible DNA sequences form, which code for various amino acids. At the same time, the number of variable positions also determines the number and position of the introduced inosines in the complementary DNA sequence. The calculation

of the heterogeneity of such a construct is shown in Fig.
5.

As, to form the double-stranded DNA, singled-stranded DNA
containing inosines is in each case hybridized with
oligomers containing A, T, G or C in each case randomly
distributed at the corresponding positions, a mixture of
double-stranded *gp120*-sequence or of part (fragment)
thereof is obtained, the nucleic acid sequences in each
case being derived from the *env* sequence (SEQ ID NO: 1 or
12) and the nucleic acid sequences in each case being
distinguished from each other in the region coding for the
V2 loop and/or in the region coding for the V3 loop. This
means that nucleic acid sequences contained in the mixture
differ from each other such that they code for a mixture of
proteins which in each case contain amino acid sequences
which are different from each other in the V2 loop and/or
in the V3 loop.

According to a preferred version of the invention, in this
way, at least $10^2$, preferably at least $10^3$ and according to
a particularly preferred version of the invention at least
$10^4$ sequence variants can be obtained at nucleic acid level
(DNA level). This DNA mixture is hereafter called the DNA
pool, and in relation to the variants of the *gp120* sequence
this is hereafter called the pool with variable *gp120*
construct (*gp120* pool).

Within the framework of the present invention, the term
sequence variants of a DNA sequence is taken to mean
molecules which have a derived nucleic acid sequence vis-à-
vis a native viral DNA molecule or a part (fragment) of

same, the variants being distinguished from each other in that at least one nucleotide can be exchanged at any chosen points of the sequence. Preferably, the sequence variants contain several nucleotide exchanges at various points of the sequence, the number and the position of the exchanged nucleic acids essentially depending on the length of the nucleic acid sequence.

Upon the expression of the *env*-gene variants starting from the produced plasmid-DNA mixture, it is to be expected that, because of the random distribution, all possible DNA sequences are expressed. Therefore, all sequence variants that are possible, on the basis of the pre-set DNA mixture, of the *gp120* molecule or of part of same are then also formed. The heterogeneity of the *gp120* mixture is calculated on the one hand using the heterogeneity of the DNA sequence and the degeneration of the genetic code for the amino acids (cf. also Fig. 5).

The heterogeneity of the plasmid-DNA mixture is demonstrated by DNA sequencing of single clones. To this end, the mixture of the plasmid DNA is transformed in *E. coli* and single clones are randomly selected and their V2 or V3 loop sequence determined. In total, approx. 100-200 different clones are to be sequenced. The statistical distribution of the DNA sequences can be use to calculate the distribution of the whole mixture. The method corresponds very largely to the method of random sampling such as is used as standard for the quality control of the most varied products.

The direct molecular analysis of the heterogeneity of the GP120 protein mixture proves to be somewhat more difficult, as single GP120 molecules cannot be separated from the mixture and identified. This is in the nature of things, as the individual GP120 variants differ from each other in a few amino acids only. Gel electrophoresis separation or mass spectroscopic analysis can establish whether a mixture or a single form of the GP120 molecule is involved. When the GP120 mixture is used e.g. in animals as a vaccine, the immune response induced in the animal is directly dependent on the number and the composition of the GP120 mixture. The immune response of the animal, the neutralizing antibodies can then be examined in virus neutralization tests. To monitor a corresponding variant-overlapping immune response, various patient isolates of HIV-1 can be used in the neutralization test. Preferably, patient isolates are chosen which differ in the ability to use various co-receptors for the infection of the target cells. The neutralization potential of the GP120 mixture then serves as a quality measure.

A subject of the present invention is therefore furthermore a protein mixture which comprises GP120 proteins which contain amino acid sequences which in each case differ from each other in the V2 loop and/or in the V3 loop, the mixture containing, according to one version of the invention, at least $10^2$, preferably at least $10^3$, according to a version of the invention, and, according to a particularly preferred version of the invention, at least $10^4$ sequence variants.

As already mentioned earlier, the mixture of double-
stranded DNA which can be obtained by hybridization of
inosines-containing single-stranded DNA with randomly
distributed A-, T-, G- and/or C-containing single-stranded
DNA, i.e. the pool with variable GP120 construct, is
transformed and fermented in prokaryotic or eukaryotic host
cells, preferably *E. Coli*.

A subject of the present invention is therefore plasmids
which contain inserted double-stranded DNA, in each case
comprising hybrids of the single-stranded, inosine-
containing nucleic acid sequence (see above) with the
single-stranded nucleic acid sequence containing a mixture
of all four nucleotide variants (A, T, G, C) (see above).
Furthermore, the invention relates to a vector mixture
which contains a mixture of these plasmids, the nucleic
acid sequences of the plasmids in each case differing from
each other in the region coding for the V3 loop and/or in
the region coding for the V2 loop. According to one version
of the invention, the vector mixture comprises at least $10^2$,
preferably at least $10^3$ and according to a particularly
preferred version of the invention at least $10^4$ of the named
plasmids. Depending on which host cells are to be used to
express the vectors (plasmids), various expression systems
and base vectors well-known to a person skilled in the art
can be considered (cf. Methods in Enzymology, Vol. 185,
Gene Expression Technology, 1991, Publ. D.V. Goeddel,
Academic Press, Inc.). Thus, for the expression in *E. coli*,
the plasmid pUC 18/19 is preferred as base plasmid into
which the nucleic acid sequence variants are cloned. Within
the framework of the present invention, the vector mixtures
can thus be expressed either in bacterial host cells such

as *E. coli*, or in eukaryotic host cells, preferably from the group consisting of Cos, CHO or baby hamster kidney cells (BHK cells) or in other host cells.

The present invention thus furthermore relates to *E. coli* host cells or eukaryotic host cells which are transfected with a vector mixture according to the invention.

As host cells are transfected with the DNA pool (the vector mixture), it is to be expected that as great or at least approximately as great a number of differently transformed host cells forms as there are sequence variants present in the DNA pool. Upon the preparation of the vaccine, special attention is to be paid to the yields during the individual cloning steps. The yields upon preparation of the double-stranded DNA fragments, the ligation of the V3 loop and V2 loop DNA fragments into the *env* gene and the transformation or the preparation of bacteria and cell clones must be such that the heterogeneity that is to be achieved of the mixture is not restricted. This is to be illustrated using an example. If, e.g. a number of balls are to be removed from a quantity of balls which differ on the basis of two colours, but there is to be a 99.9% probability that both colours are present in this selection, approx. 13 balls would have to be removed (G. Schreiber, Ein kombinatorisches Problem aus der Genetik; Bio-engineering 1988, 2: 32-35). Transferred to the cloning steps of the HIV vaccine, this means: If a heterogeneity of approx. 6000 virus variants is to be produced, 13 times as many clones would have to be produced in each individual cloning step. Thus, a gene bank of approx. 80,000 clones must be prepared for the V3 loop. Starting from this gene bank, the

expression vectors for the transfection of the CHO cells are then prepared. During transfection of the CHO cells, consequently, approx. 80,000 transfections must then be achieved. Such a mixture of transfected CHO cells then produces the desired quantity of 6000 different gp120 variants.

The present invention also relates to a process for the preparation of a nucleic acid sequence coding for a viral protein (expression cassette) in which the sequence is so modified, or preferably so many silent mutations are introduced, that it then contains at least two and preferably exclusively monovalent restriction sites. Particularly preferably, the sequence still contains only monovalent restriction sites. Preferably, the protein coded by the nucleic acid sequence is GP120, the *env* wild-type sequence coding for the GP120 being varied by silent mutations.

A further subject of the present invention is a process for the preparation of the vector mixture according to the invention, which preferably contains a mixture of plasmids, the nucleic acid sequences of which differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop, the plasmids according to the invention being ligated into a (base) vector which can be expressed in host cells (preferably *E. coli*, Cos, CHO or BHK cells). The (base) vector is preferably the pUC 18, pUC 19 or BSCenvATG vector.

Within the framework of the present invention, a process is furthermore made available for the preparation/production

of host cells, preferably chosen from the group consisting
of *E. coli*, Cos, BHK or CHO cells, in which the host cells
are transformed with a vector mixture according to the
invention, which contains a mixture of plasmids the nucleic
acid sequences of which in each case differ from each other
in the region coding for the V2 loop and/or in the region
coding for the V3 loop.

Finally, according to the invention, a process is for the
first time made available for the preparation of a DNA
vaccine, in which the process according to the invention
for the preparation of the vector mixture is carried out,
the plasmids according to the invention expressing the
mixture of the *gp120* proteins after application in host
cells (e.g. human and animal monocytes). The DNA vaccine
can optionally be formulated with pharmaceutically
compatible auxiliaries and/or carriers for application and
after administration in the organism, make possible the
production of the sequence variants of viral proteins.

The present invention thus also relates to a pharmaceutical
composition or a DNA vaccine which codes for a mixture of
structurally different virus proteins, the vaccine
containing a mixture of sequence variants of a viral DNA
molecule or of part of same, i.e. of DNA molecules the
nucleic acid sequences of which differ from each other in
the region coding for the protein, a part or a fragment of
same. By the term "structurally different virus proteins"
is meant according to the invention those proteins the
amino acid sequences of which are derived from the wild-
type sequence of the corresponding virus protein, the amino
acid sequences deviating from each other insofar as they

differ from each other and compared with the wild-type
sequence through one or more exchanged amino acids at
identical or different positions of the sequence. As
already mentioned, the nucleic acid sequences in the
vaccine code preferable according to the invention for a
mixture of structurally different GP120 proteins of HIV
(sequence variants of GP120), a vaccine being particularly
preferred which contains a mixture of DNA molecules, the
nucleic acid sequences of which differ from each other in
the region coding for the V2 loop and/or in the region
coding for the V3 loop of *gp120* of the HIV-1. In this
context, reference is also made to Fig. 3 in which known
structural differences or variations in the V3 loop of
GP120 are shown.

The DNA vaccine according to the invention is of particular
importance within the framework of a gene therapy.

Finally, according to the invention, a process for the
preparation of a pharmaceutical composition or a protein
vaccine is for the first time made available in which the
host cells according to the invention, i.e. host cells
which are transformed with a vector mixture according to
the invention, the vectors in each case containing
plasmids, the nucleic acid sequences of which differ from
each other in the region coding for the V2 loop and/or in
the region coding for the V3 loop, are cultivated under
conditions which allow the expression of the mixture of
viral protein sequence variants. The host cells are
preferably bacterial host cells, such as *E. coli*, or
eukaryotic host cells, preferably from the group consisting

of Cos, Chinese hamster ovary (CHO) or baby hamster kidney cells (BHK cells).

Through the present invention, it is thus possible to make available a vaccine comprising variable GP120 proteins which, in the prophylactic and/or therapeutic treatment of a HIV infection or AIDS

1.  activates the immune system,
2.  induces the formation of HIV-1-neutralizing antibodies,
3.  prevents the loss of neutralizing antibodies and
4.  in which monitoring of the stimulation of the GP120-specific immune response is taken over by the active ingredient.

This is preferably carried out in that, even before the loss of HIV-neutralizing antibodies associated with the outbreak of AIDS, sufficient antigens are available which, because of their variety of different amino acid sequences (i.e. their sequence variations) are in a position to induce the formation of those HIV antibodies which normally, i.e. without corresponding prophylaxis according to the present invention, are lost or experience a reduction in concentration as the illness progress. In addition to a preventative treatment, the present invention also relates to therapeutic treatment, the GP120 mixture (protein vaccine) being used, to counter a reduction or a loss of HIV-neutralizing antibodies by activation of the immune system and induction of the formation of new or additional HIV-1-neutralizing antibodies.

According to the invention, the mixture of structurally different viral proteins which are sequence variants of a viral protein (preferably of GP120) or of part of same is therefore used for the preparation of a vaccine for the prevention and or therapy of a viral infection in humans. Furthermore, the present invention relates to the use of a mixture of DNA molecules which code for sequence variants of a viral protein or of part of same for the preparation of a vaccine for the prevention and/or therapy of a viral infection in humans. According to a preferred version, the invention relates to the use for the preparation of a vaccine for the prevention and/or therapy of a HIV infection in humans.

Within the framework of the present invention, the virus infection can be any infection in which self-replicating virus variants form in the course of the illness, where those amino acid sequence sections or nucleic acid sections coding for the latter can be considered according to the invention as selected protein sequences or protein-coding nucleic acid sequences, in which, in the course of viral illnesses, variants, i.e. sequence variations, are always or often observed. Preferably, in the present case, these are sequences variants of the GP120 molecule (at protein level) or of the region coding for GP120 (at DNA level) or parts of same, in particular sequence variants in the V3 loop or in the V2 loop, preferably both in the V3 and in the V2 loop.

According to the invention, pharmaceutical compositions or vaccines are consequently made available which, in the immunoreconstitutive treatment of those suffering viral

infection, activate the immune system in such a manner that the naturally acquired immune defence is regenerated against the virus and restimulated to thus prevent the virus variants replicating in the patient from multiplying. According to a preferred version of the invention, vaccines for the immunoreconsitutive treatment of those infected by HIV-1 are made available for the first time.

Within the framework of the present invention, the vaccines can either be formulated or administered per se, i.e. as DNA and/or protein mixtures without further additives, or together with further active ingredients, such as e.g. immunostimulants such as interleukin-2, CC and CXC-Chemokine, and/or pharmaceutically compatible auxiliaries and carriers. As a rule, the vaccines according to the invention are formulated for intravenous application, such as e.g. by the intravenous, intramuscular, subcutaneous routes. In addition, the vaccines can be formulated for oral, mucosal (intravaginal, intrarectal) and transdermal application.

The present invention has the following advantages:

Through a selective genetic manipulation of the nucleic acid sequence, with the help of a gene cassette, any desired variants of a pathogen or of a gene of the exciter can be prepared. The variations preferably affect the regions against which the neutralizing antibodies or cytotoxic T cells are formed. An inoculant in the form of a mixture of variants of the pathogen or of variants of the corresponding antigens of a pathogen can thereby be prepared. The advantages of this concept in relation to HIV

as a pathogen lie in the preparation of a mixture of HIV-GP120, the external membrane protein of HIV, against which the virus-neutralizing immune response of humans is directed. As, in the precise case of the HIV illness, each patient is exposed to a variety of HIV variants which all differ in the sequence of the GP120 protein, the use of a variety of GP120 variants as inoculant is of particular advantage, this GP120 mixture being able to be used both for the immunization of normally healthy persons and for the therapy of persons already infected by HIV. In the case of immunization with the GP120 mixture, as broad as possible an immune response against as many virus variants as possible in humans is induced, which in the ideal case protects against all HIV variants. In the case of therapy with the GP120 mixture, the loss of neutralizing antibodies and the loss of the HIV-specific cellular immune response can be combated.

According to the invention, it is furthermore possible for the first time to prepare *gp120* clones which, in the nature or variety of the observed sequence variations, correspond to plasma isolates of those infected with HIV or ill with AIDS (cf. M. Schreiber et al., J. Virol. 68 No. 6 (1994) 3908-3916). In this context, in contrast to various approaches followed to date, no isolated GP120 molecule or an antigenic fragment of same is used for immunization, but a protein mixture which consists of a variety of sequence variants which are characterized by a complete GP120 sequence which in addition contain the correct tertiary structure corresponding to the natural plication of the GP120 molecule. The conformation of the virus variants made available in the vaccine is important insofar as sequence

and structure variants are thereby provided which, with the GP120 variants detectable in the course of the viral illness, as great as possible a similarity with regard to the observed sequence variations and also the conformation required for the bonding to CD4 and the co-receptor is achieved, whereby an effective immune stimulation can be achieved.

Through the invention, a mixture of GP120 proteins is provided which represents the protein vaccine. At the same time, a mixture of the genes coding for this protein mixture is provided through the invention. These genes can be transferred into a vector which is suitable for direct application in humans (DNA vaccine) (Ulmer et al., 1995 Ann NY Acad Sci 772:117-125; Donnelly et al., 1995 Ann NY Acad Sci 1995 772:40-46). A DNA vaccine has the advantage that the heterogeneity of the mixture of the DNA vectors can be greater than that of the mixture of the recombinant GP120 proteins. It is easier to produce a high heterogeneity of a DNA vector mixture. Such a DNA vaccine would cover a much broader spectrum of HIV variants. From the technical point of view, the preparation of the DNA mixture is easier. A precondition for the DNA vaccine is a *gp120* expression vector which carries the two cleavage sites BstEII and BamHI. This makes possible the conversion of the V3 loop and V2 loop variables *env* gene fragments into such a vector.

With regard to the GP120 proteins, it is to be borne in mind that 5 potential points for N-glycosylation are located in the region of the V3 loop. The sugar groups protect the V3 loop against recognition by neutralizing

antibodies. Virus variants which contain a glycosylated V3 loop are more poorly neutralizable than virus variants in which glycosylation is incomplete. As virus variants with a complete V3 loop glycosylation can escape the neutralizing immune response, they are important for the transfer and for the establishment of the infection. In the course of the illness, if a part of the neutralizing antibody response is lost, virus variants whose V3 loop is not completely glycosylated establish themselves. As the V3 loop is no longer covered by sugar groups, these virus variants replicate faster than the completely glycosylated V3 loop mutants. According to the invention, it is therefore advantageous to take the varying glycosylation of the GP120 V3 loop into account when constructing the GP120 vaccine.

Within the framework of the present invention, the principle represented with reference to HIV can be transferred to other viruses which also form variants in the course of the viral illness. In particular, vaccines can be provided against a variety of viruses which induce a broad immune response against many virus variants in humans which in the ideal case protects against all variants. In the case of therapy with such vaccines, the loss of neutralizing antibodies and the loss of the HIV-specific cellular immune response can be effectively combated. In this context, it is furthermore possible or even advisable, with pathogen types (viruses) which form several sequence variants of several proteins in the course of an illness, to provide several gene cassettes according to the invention each of which contains the nucleic acid sequence coding for variants of each of these proteins or fragments

of same, in order to combat the loss of neutralizing antibodies against all conceivable virus variants.

Within the framework of the present invention, the advantages of a protein vaccine and a DNA vaccine can be advantageously combined in order to increase the success of a preventative and/or therapeutic treatment of a viral illness. A pharmaceutical composition is thus furthermore provided for the prevention and/or therapy of a viral infection, which comprises a protein mixture and a nucleic acid mixture, the protein mixture comprising sequence variants of a viral protein or of part of same, and the nucleic acid mixture comprises DNA molecules which code for sequence variants of a viral protein or of part of same. In particular, the pharmaceutical composition is a combination preparation which comprises both a protein mixture comprising one of the above-named sequence variants of the GP120 protein and a nucleic acid mixture derived from an above-named *env* sequence in SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same.

The present invention is explained in the following, using examples, figures and a sequence protocol.

EXAMPLES

1.  General description of the preparation process

    1.1  Cloning of V3 loop-coding oligonucleotides
    1.2  Cloning of the V3 *env* variants
    1.3  Analysis of variability

2.  Material and methods

    2.1  Abbreviations
         2.1.1     General abbreviations
         2.1.2     Nucleic acids
    2.2  Bacterial strains
    2.3  Plasmids
    2.4  Enzymes
    2.5  Chemicals
    2.6  Oligonucleotides
    2.7  Molecular weight standards
    2.8  Reagent kits used
    2.9  Media
    2.10 Sterilization of solutions
    2.11 Cultivation and storage of bacteria
    2.12 Preparation of competent cells
    2.13 Transformation into E. coli
    2.14 Plasmid-DNA preparation
    2.15 Separation of DNA in agarose gels
    2.16 Purification of DNA from agarose gels
    2.17 Separation of DNA in polyacrylamide gels
    2.18 Cleaving of DNA
    2.19 Ligation of DNA
    2.20 DNA sequencing

## 1.  General description of the preparation process

### 1.1  Cloning of a mixture of V3 loop *gp120* variants

The oligonucleotides are synthesized as described under
2.6. As V3 loop sequence, a mutated sequence of the HIV-1
patient isolate F1-01 (M. Schreiber et al., J. Virol. 68
no. 6 (1994) 3908-3916) is given by way of example. Any
other sequence or mixture of sequences is also possible. To
clone many different variants of a V3 loop sequence,
mixtures are used instead of pure nucleotide components at
specific positions of the sequence. In this way, a mixture
is obtained of oligonucleotides which all differ in the
sequence and thus code for different V3 loops. Such
oligonucleotide mixtures are also called oligonucleotides
with degenerated sequences. Starting from chemically
synthesized oligonucleotides with degenerated sequences,
the region coding for the various V3 loops is shown.

An oligonucleotide, in reading-frame orientation (forward),
is synthesized for the *env*-sequence section from the BglII
cleavage site to the first degenerated position. A second
oligonucleotide, in complementary orientation (reverse), is
synthesized according to the sequence from a position which
lies 15 bases before the beginning of the variable region
(method 2.6). Through overlapping of the 3' region along a

length of 15 bases in total, the hybridization of both oligonucleotides is carried out. In a subsequent reaction with DNA polymerase (e.g. Taq DNA polymerase or Klenow fragment), a completely double-stranded DNA molecule forms from the two hybridized oligonucleotides (method 2.21).

The thus-obtained DNA mixture is digested with restriction endonucleases BglII and XbaI. The cloning of the DNA mixture is carried out in the expression vector ΔV3 pBSCenvV3 cleaved with BglII and XbaI (methods 2.15 and 2.16).

This vector contains the coding sequence of the *gp160* of the HIV-1 strain NL4-3 (IIIB). The NL4-3 *env* gene was manipulated such that it possesses the restriction cleavage sites BglII and XbaI as well as ApaLI, PstI and BclI once each only. The region coding for the V3 loop, which lies between the cleavage sites BglII and XbaI, was removed and replaced by a 15 base pair sequence, whereby an analytical cleavage site was introduced for the enzyme AscI. The BglII- and SbaI- cleaved V3 loop DNA mixture is cloned into this vector (method 2.19). The loss of the AscI cleavage site in the finished V3 loop pBSCenvV3 vector can be used for the selection of the V3 loop coding clones (method 2.18).

The thus-formed plasmid mixture is transformed in DH5α bacteria (method 2.12), a transformation rate of >$10^5$ having to be achieved. A gene library of V3 loop fragments with a size of approx. $10^5$ clones is thereby obtained. The mixture is to be analyzed by DNA sequencing (see 1.3). With the help of this process, a mixture of clones is obtained,

which all code for different V3 loop variants of the GP120
protein. This mixture of vectors serves as starting product
for the preparation of the protein mixture for use as
inoculant and as starting product for use as DNA vaccine.

For the preparation of the GP120 protein mixture, the
pBSCenvV3 expression vectors are transfected into COS cells
and CHO cells (method 2.22). The purification of the
different GP120 proteins, the active ingredient mixture, is
carried out according to method 2.23, as described in the
literature.

## 1.2 Cloning of a mixture of V2 loop *gp120* variants

Variants of the V1 loop and V2 loop are prepared in the
same way. The expression vector ΔV3-pBSCenvV3 has three
additional restriction cleavage sites for this. The
variation of the V1 loop takes place through cloning into
the ApaLI and PstI cleavage sites. The variation of the V2
loop takes place through cloning into the PstI and Bc1I
cleavage sites.

## 1.3 Analysis of variability

The analysis of the V3 loop mixtures takes place through
DNA sequencing (method 2.20). The statistical selection of
approx. 100-200 clones whose V1, V2 and V3 loop sequences
are to be determined, is planned. If all these clones are
different, the heterogeneity of the gene library and thus
also the heterogeneity of the protein mixture can be
measured with the help of a statistic calculation.

## 2. Material and methods

## 2.1 Abbreviations

### 2.1.1 General abbreviations

μg   microgram

μl   microlitre

μmol micromole

APS  ammonium persulphate

bp  base pairs

dNTP desoxynucleoside triphosphate

DNA  desoxyribonucleic acid

DTT  dithiothreitol

EDTA ethylenediaminetetraacetic acid

g   gram

GP  glycoprotein

h   hour

HIV  human immunodeficiency virus

IPTG isopropyl-β-D-thiogalactopyranoside

MOPS 3-morpholinepropanesulfonic acid

OD  optical density

PAGE polyacrylamide gel electrophoresis

PCR  polymerase chain reaction

RT  room temperature

TEMED N,N,N',N'-tetramethylethylenediamine

### 2.1.2 Nucleic acids

A   adenine

C   cytosine

G   guanine

T   thymidine

## 2.2  Bacterial strains

Escherichia coli DH5α:   F-, endA1, hsdr17,
(rk-mk+), su-pE44, recA1,
λ-, gyrA96, relA1, Φ80d lac
z ΔM15

## 2.3  Plasmids

pBSCenvATG      own design, the sequence is given in
SEQ ID NO: 10.

## 2.4  Enzymes

Restriction enzymes MBI-Fermentas, Gibco-BRL, Biolabs

DNA polymerases       MBI-Fermentas, Gibco-BRL

T4-DNA ligase         MBI-Fermentas

## 2.5 Chemicals

[α-35S]dATP                 Amersham Life Science

Agarose, ultra pure         GIBCO BRL

Ammonium persulphate        Merck

Ampicillin                  US Biochemical

Bacto-Agar                  Becton Dickinson

Bacto-Trypton               Becton Dickinson

Boric acid                  Merck

Bromphenol blue             Merck

Calcium chloride            Merck

| | |
|---|---|
| Desoxyribonucleotides | MBI-Fermentas |
| Dithiothreitol | Biotechnik, St. Leon-Rot |
| Glacial acetic acid | Merck |
| Ethidium bromide | Sigma |
| Ethylenediaminetetraacetic acid | Merck |
| Glycerine | Merck |
| Urea | ICN Biomedicals |
| Yeast extract | Becton Dickinson |
| Potassium chloride | Merck |
| Potassium dihydrogen phosphate | Merck |
| Magnesium chloride | Merck |
| Acrylamide mix | Roth |
| Sodium acetate | Merck |
| Sodium chloride | Merck |
| Disodium hydrogen phosphate | Merck |
| Sodium dihydrogen phosphate | Merck |
| 2-propanol | Merck |
| Sigmacote (chlorinated polysiloxane) | Sigma |
| N,N,N',N'-tetramethylethylenediamine | Merck |
| Tris(hydroxymethyl)-aminomethane | GIBCO BRL |

## 2.6  Oligonucleotides

All oligonucleotides listed were prepared with the Expedite™ nucleic acid synthesis system of PE Biosystems (Weiterstadt). Oligonucleotides are used for the cloning of

*env* genes which differ from each other only in the sequence for the V3 loop. The sequences of the oligonucleotides are listed taking as an example the cloning of the V3 loop region for the HIV-1 patient isolate F1-01.

V3 loop: for cloning into pNL4-3/Bg1II-NheI, F1-01, forward:

5'-AAG ATG TAG TAA TTA GAT CTG CCA ATT TCA CAG ACA ATG CTA AAA CCA TAA TAG TAC AGC TGA ACA CAT CGT TAG AAA TTA ATT GTA CAA GAC CCA ACA ACA AT ACA-3' (SEQ ID NO: 3)

V3 loop: for cloning in pNL4-3/Bg1II-NheI, F1-01:

5'-TTT TGC TCT AGA AAT GTT ACA ATG TGC TTG TCT TAT GTC TCC TGT TGC AGC TTC TGT TGC ATG AAA TGC TCT CCC TGG TCC GAT ATG GAT ACT ATG-(GA)(AT)(GATC) TTT TCT TGT ATT GTT GTT GGG-3' (SEQ ID NO: 4)

For sequencing of V3 loop clones, the oligonucleotides 7010, 7011 and the M13 standard primer were used (M13, M13r).

Primer 7010: 5'-CCA TGT ACA AAT GTC AG-3' (SEQ ID NO: 5)
Primer 7011: 5'-AAA ACT GTG CGT TAC AA-3' (SEQ ID NO: 6)
M13-Primer:  5'-GTA AAA CGA CGG CCA GT-3' (SEQ ID NO: 7)
M13r-Primer: 5'-CAG GAA ACA GCT ATG AC-3' (SEQ ID NO: 8)

## 2.7  Molecular weight standards

1 kb conductor       MBI-Fermentas
100 bp conductor     MBI Fermentas

## 2.8 Reagent kits used

| | |
|---|---|
| T7 Sequencing-Kit | Pharmacia |
| Qiaquick PCR purification Kit | Qiagen |
| Qiagen Plasmid Kit | Qiagen |

## 2.9 Media

YT Medium:

> 10g trypton
> 5g yeast extract
> 5g NaCl

dYT medium:

> 16g trypton
> 10g yeast extract
> 5g NaCl

dYT agar plates:

> 10g trypton
> 5g yeast extract
> 5g NaCl
> 15g agar

The quoted quantities relate in each case to 1000 ml deionized water. The mixtures were autoclaved at 121°C and 1.5 bar for 20 mins. Should the media contain antibiotics or other heat-sensitive reagents, the corresponding quantities were sterile-filtered and added after cooling of the medium.

> Ampicillin     3.3 ml/l (60 mg/ml)

```
IPTG              3 ml/l (100 mM)
xgal              3 ml/l (2% in DMF, do not filter)
```

## 2.10 Sterilization of solutions and apparatus

Solutions and media as well as pipette tips, Eppendorf
vessels and glass apparatus were autoclaved for 20-40 mins
at 121°C and 1.5 bar. Heat-sensitive solutions such as e.g.
antibiotics and IPTG solutions were sterile-filtered.

## 2.11 Cultivation and storage of bacteria

Bacteria cultures were in each case inoculated with a
single colony. To isolate a single colony, cells of a
liquid culture were spread on an agar plate. After
incubation at 37°C overnight, isolated colonies have grown.
For the short-term preservation of the bacteria, the agar
plates were sealed with parafilm and stored at 4°C. For a
long-term storage of bacteria strains, 0.75 ml of a YT-
overnight culture were mixed with 0.25 ml sterile
glycerine, shock-frozen in liquid nitrogen and stored at -
70°C.

## 2.12 Preparation of competent cells

100 ml YT-medium were inoculated with 100 µl of an
overnight culture. The bacteria were incubated until an
$OD_{580}=0.4$ at 37°C was reached in the shaking incubator and
then centrifuged for 8 mins at 4°C and 1000g. All
subsequent work was carried out on ice using pre-cooled
vessels and 4°C cold solutions. The cells were resuspended
in 50 ml 50 mM $CaCl_2$ and incubated for 30 mins on ice. After

renewed centrifuging, the bacteria were taken up in 2.5 ml
sterile TFBII buffer and divided into portions of 100 µl
each. The 100-µl portions of the competent cells were able
to be, insofar as they were not required for an immediate
transformation, stored in liquid nitrogen at -70°C after
shock-freezing.

TFBII buffer:

| | | |
|---|---|---|
| 10 mM | MOPS pH 7.0 |
| 75 mM | CaCl$_2$ |
| 10 mM | KCl |
| 15% | glycerine |

## 2.13 Transformation of *E. coli*

(Hanahan, J. Mol. Biol. 166 (1983) 557-580)
A 100 µl portion of competent cells was thawed in an ice
bath and incubated with 1-100 ng plasmid-DNA for 30 mins at
0°C. The transformation mixture was then incubated for 1
min at 42°C. 1 ml YT-medium was then added and agitated for
an hour at 37°C. To make possible a selection of the
transformed cells, 100-500 µl of the mixture were spread
onto antibiotic-containing YT agar plates. After incubation
overnight at 37°C, only colonies which carry the plasmid-
coded resistance gene have grown. Upon a transformation of
cells which do not contain any functional β-galactosidase
(lacZΔM15 mutation, e.g. DH5α), vectors such as the pUC
used make possible a direct selection (*blue white
screening)* of recombinant bacteria via a plasmid-coded β-
galactosidase. For the blue-white selection, the bacteria
were spread onto Amp/IPTG/Xgal-YT agar plates. The
substrate Xgal is cleaved into a blue dye by the β-

galactosidase. Because of the destruction of the reading frame of the lacZ gene by the insertion of a foreign DNA fragment into the lacZ gene, white colonies are produced. On the other hand, blue colonies mean that the lacZ gene has remained functional during cloning and ligation and no DNA was inserted.

## 2.14 Plasmid DNA preparation

(Qiagen, Hilden)

To isolate plasmid-DNA from *E. coli*, the QIAprep Spin Miniprep Kit from Qiagen was used. The DNA preparation took place according to the manufacturer's instructions (QIAprep Plasmid Handbook 03/95). Plasmid-containing *E. coli* bacteria were inoculated in dYT-amp medium and agitated overnight at 37°C. Three ml of the *E.coli* overnight culture were used for the plasmid isolation. The bacteria were harvested (1 min, 14,000 rpm, Heraeus centrifuge) and taken up in 250 µl P1 buffer. The bacteria were decomposed through alkaline lysis (250 µl, 0.2N NaOH/1%-SDS). The mixture was neutralized through the addition of 3M potassium acetate, (350 µl, pH 5.5). The purification of the plasmid DNA is based on the selective bonding of plasmid DNA to DEAE anion exchanger columns. With the chosen salt concentrations and pH conditions, the plasmid DNA bonds to the DEAE matrix. The DEAE matrix was washed (750 µl, PE buffer, Qiagen). The plasmid DNA was then eluted with 50 µl $H_2O$ and stored at -20°C.

To obtain plasmid DNA in a larger quantity (approx. 100 µg), 25 ml overnight culture were used. To purify the DNA, the QIAGEN Plasmid Midi Kit was used. The purification is

based in this case on the principle described under the "QIAprep Spin Miniprep Kit" and was carried out according to the manufacturer's instructions (QIAGEN Plasmid Purification Handbook 01/97).

## 2.15 Separation of DNA in agarose gels

The separation of DNA took place through gel electrophoresis in agarose gels. Depending on the size of the fragments examined, 0.8-2% agarose were dissolved by boiling-up in TBE buffer (for analytical agarose gels) or TAE buffer (for preparative agarose gels). After cooling to approx. 60°C, the gel liquid was mixed with 1 µl ethidium bromide (10 mg/ml) and poured into a prepared gel bed with inserted comb. The completely cooled gel was coated in an electrophoresis chamber with TBE or TAE buffer and the comb removed. The DNA samples were mixed with 1/5 vol. coating buffer and pipetted into the sample bags. The fragments were separated at 5-10 volt/cm gel length for 0.5-2 h. For detection, the gel was photographed after the electrophoresis in UV transmitted light. The molecular weights of the DNA bands were measured compared with following DNA molecular weight markers.

TBE buffer:

|        |            |
|--------|------------|
| 100 mM | Tris       |
| 100 mM | boric acid |
| 3 mM   | EDTA       |

TAE buffer:

|       |      |
|-------|------|
| 40 mM | Tris |
| 2mM   | EDTA |

0.114%    glacial acetic acid

Coating buffer:

0.25%     bromphenol blue

0.25%     xylene cyanol

30%       glycerine

50 mM     EDTA

## 2.16 Purification of DNA from agarose gels

For the extraction of the DNA from agarose gels, the
"QIAquick Gel Extraction Kit" (Qiagen, Hilden) was used.
The buffer conditions were selected such that the nucleic
acids bond to the silica membrane of the columns, whilst
low-molecular-weight bacteria components pass through the
membrane (method 2.15). The process was carried out
according to the "Qiaquick Gel Extraction Kit Protocol" of
the QIAquick Spin Handbook 07/97. The purified DNA was in
each case eluted from the silica membrane with 50 µl $H_2O$.

## 2.17 Separation of DNA in polyacrylamide gels

The separation of radioactively marked DNA fragments for
the DNA sequencing was carried out using denaturizing
polyacrylamide gels. Two purified glass plates were cleaned
of fatty residues with ethanol and coated with 1 ml
Sigmacote per plate to obtain a hydrophobic surface.
Through the coating, a later tearing of the gel upon
removal from the glass plate was to be avoided. Spacers
were placed between the two glass plates, and served at the
same time for lateral sealing. The whole apparatus was
fixed with several clamps. To prepare the sequence gel, 21

g urea were dissolved in approx. 20 ml water. After the addition of 7.5 ml acrylamide mix and 5 ml 10x TBE buffer (see 2.15), the mixture was made up to 50 ml with water. The polymerization of the gel was started by the addition of 300 µl APS and 100 µl TEMED. Immediately afterwards, the gel solution was poured into the prepared gel apparatus and the pocket bottom shaped by inserting the flat back of the saw-tooth comb. The gel was stored horizontally at room temperature until polymerization was complete. After the insertion of the gel into the gel chamber, the latter was filled with TBE buffer. By rotating the saw-tooth comb, the deposit pockets were formed. A 20-min pre-run was carried out in order to heat the gel to the optimal operating temperature. The pockets were washed thoroughly with TBE buffer and then filled with 4-5 µl of the samples. The electrophoresis took place at 2 kV (150 watt) for approx. 2-3 h. The gel was removed from the gel chamber after the electrophoresis was ended and one of the glass plates raised carefully. By laying-on and light pressing, the gel was transferred onto Schleicher & Schuell paper (Whatman, England). After drying at 80°C under vacuum, the gel was exposed on an X-ray film (Kodak BioMax MR-1, Sigma Deisenhofen) for 1-3 days at RT.

Acrylamide mix:

     40% polyacrylamide

     0.8% bisacrylamide

## 2.18 Cleaving of DNA

Restriction endonucleases recognize and hydrolyze enzyme-specific palindromic DNA sequences of mostly 4-8

nucleotides' length. Upon hydrolysis, depending on the type of enzyme, blunt ends or overhanging ends of single-stranded DNA form. For a restriction digestion, 0.1-60 µg DNA were incubated for 2-5 h at 37°C with 2-120 units of a restriction enzyme in the buffer indicated by the manufacturer. The total volume was between 10 µl for analytical and 300 µl for preparative mixtures. For restrictions with two different enzymes, a buffer was chosen in which both enzymes have a sufficient activity, e.g. the EcoRI buffer for an EcoRI/BamHI digestion, or the conditions for the second enzyme were set after incubation with the first enzyme.

## 2.19 Ligation of DNA

DNA ligases catalyze the linking of DNA molecules, consuming $NAD^+$ or ATP through formation of a phosphodiester bond between a free 5' phosphate group and a 3' hydroxyl group. A three- to five-times excess of DNA fragment was incubated overnight at 12°C in ligation buffer with 100-500 ng cleaved plasmid and 5 units T4-DNA ligase as well as 10 nmol ATP. Only ligations of cohesive ends were carried out.

Ligation buffer:

| | |
|---|---|
| 40 mM | Tris-HCl pH 7.8 |
| 10 mM | $MgCl_2$ |
| 10 mM | DTT |
| 0.5 mM | ATP |

## 2.20 DNA sequencing

(Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977) 5463-5467) The sequencing of DNA was carried out according to the chain-break method. Double-stranded plasmid DNA serves as substrate. Double-stranded DNA can be synthesized from single-stranded DNA starting from an oligonucleotide primer in the presence of dNTPs through a DNA polymerase. If a small proportion of didesoxynucleotides (ddNTPs) is contained in the nucleotide mixture, this leads to a randomly distributed incorporation of the ddNTPs, as the DNA polymerase cannot distinguish between dNTPs and ddNTPs. The incorporation leads to a breaking of the double-stranded synthesis, because of the absence of the 3'-hydroxyl group of the ddNTPs and, as it takes place statistically distributed, to DNA single strands of different lengths. As the synthesis mixture is divided in four and only one of the four ddNTPs is present in each sample, base-specific chain breaks result in the respective mixtures. To mark the synthesized single strand, $\alpha$-$^{35}$S-dATP is added to the reaction. After denaturing and separation of the DNA via polyacrylamide gel electrophoresis, the bands can be autoradiographically detected and the sequence of the DNA strand read off direct.

The sequencings were carried out with the T7-Sequencing-Kit (Pharmacia) according to the manufacturer's instructions. Either the M13-Universal-Primer (Pharmacia) or Primer M13r was used as oligonucleotide primer. 32 μl (approx. 2 μg) of purified double-stranded plasmid DNA were denatured with 8 μl 2 M NaOH for 10 mins at RT. After the addition of 7 μl 3 M sodium acetate, pH 4.8, 4 μl H2O and 120 μl -20° ethanol, the DNA was precipitated for 20 mins at -70°C. The DNA was isolated by centrifuging (15 mins, 4°C), washed twice with

-20° cold 70% ethanol and dried in a vacuum centrifuge. The DNA sediment was then resuspended in 10 µl H$_2$O and after the addition of 2 µl *annealing* buffer and 2 µl primer solution (10 pmol in water) hybridized with the oligonucleotide primer at 65°C for 5 mins, 37°C for 10 mins and 5 mins at RT. Directly afterwards, for the purpose of primer elongation and radioactive marking, 3 µl *labelling mix*, 1 µl α-$^{35}$S-dATP (1000 Ci/mmol, 10 µCi/µl) and 2 µl T7 polymerase solution (diluted 1:5 with *enzyme dilution* buffer) were added and incubated for 5 mins at RT. In each case, 4.5 µl of this mixture were placed onto a *MicroSample Plate* (Greiner) preheated to 37°C in which 2.5 µl each of the four different dNTP/ddNTP mixes were present. After five minutes' incubation at 37°C, which served for further elongation and base-specific termination, the reactions were ended by the addition of 5 µl stop solution. Before application to the polyacrylamide gel, the samples were denatured for 2 mins at 80°C. The samples were stored at - 20°C.

Annealing buffer:

| 1 M | Tris-HCl pH 7.6 |
| 100 mM | MgCl$_2$ |
| 160 mM | DTT |

Labelling mix:

| 1.375 µM | dCTP |
| 1.375 µM | dGTP |
| 1.375 µM | dTTP |
| 333.5 mM | NaCl |

Enzyme dilution buffer:

|            |                       |
|------------|-----------------------|
| 20 mM      | Tris-HCl pH 7.5       |
| 5 mM       | DTT                   |
| 100 µg/ml  | BSA                   |
| 5%         | glycerine             |

Stop solution:

|         |                   |
|---------|-------------------|
| 10 mM   | EDTA              |
| 97.5%   | formamide         |
| 0.3%    | bromphenol blue   |
| 0.3%    | xylene cyanol     |

| A-mix: | 840 µM dCTP           | C-Mix: | 840 µM dATP           |
|--------|-----------------------|--------|-----------------------|
|        | 840 µM dGTP           |        | 840 µM dGTP           |
|        | 840 µM dTTP           |        | 840 µM dTTP           |
|        | 93.5 µM dATP          |        | 93.5 µM dCTP          |
|        | 14 µM ddATP           |        | 14 µM ddCTP           |
|        | 40 mM Tris-HCl pH 7.6 |        | 40 mM Tris-HCl pH 7.6 |
|        | 50 mM NaCl            |        | 50 mM NaCl            |

| G-Mix: | 840 µM dATP           | T-Mix: | 840 µM dATP           |
|--------|-----------------------|--------|-----------------------|
|        | 840 µM dCTP           |        | 840 µM dCTP           |
|        | 840 µM dTTP           |        | 840 µM dGTP           |
|        | 93.5 µM dGTP          |        | 93.5 µM dTTP          |
|        | 14 µM ddGTP           |        | 14 µM ddTTP           |
|        | 40 mM Tris-HCl pH 7.6 |        | 40 mM Tris-HCl pH 7.6 |
|        | 50 mM NaCl            |        | 50 mM NaCl            |

## 2.21 Preparation of double-stranded DNA using oligonucleotides

The oligonucleotides were hybridized at 60°C for the preparation of the DNA mixture. 100 pmol were used per

nucleotide. 1 pmol of the hybridized sample was used for the PCR or 100 pmol for the Klenow reaction.

The following cycles were carried out for the PCR:

| | | |
|---|---|---|
| At the start: | 10 min | 94°C |
| 30 cycles: | 1 min | 94°C |
| | 1 min | 45°C |
| | 1 min | 72°C |
| At the end: | 10 min | 72°C |

1 pmol hybridized oligonucleotide was used as DNA *template* for the PCR. PCR primers were used for all PCR reactions carried out in a final concentration of 0.1 pmol/µl. 1 unit of the Taq polymerase was used per 50 µl of the reaction mixture. The concentration of the dNTPs in the PCR mixture was set at 0.1 mM.

To prepare the DNA mixture with the help of the Klenow reaction, 100 pmol of the hybridized oligonucleotides was dissolved in Klenow buffer (50 mM Tris-HCl, pH 8.0; 50 mM MgCl$_2$, 10 mM DTT, 0.05 mM dNTP). The reaction was started by the addition of 5 units DNA polymerase I (Klenow fragment). After 30 min at 37°C, the reaction was stopped by heating to 75°C (10 min).

## 2.22 Transfection of COS cells and CHO cells

5-10 µg linearized plasmid DNA are dissolved in 150 µl cell culture medium (without FKS and antibiotics). 30 µl of the transfection agent (SuperFect, Qiagen, Hilden) are added to the solution. The mixture is incubated for 10 min at RT and added to the cells after the addition of 1 ml medium. The

COS cells (80% confluent growth) were cultivated in 60 mm plates and washed with PBS shortly before the transfection. After an incubation of 3 h at 37°C and 5% $CO_2$, the transfection agent was removed. The cells were then washed 4 times with PBS and taken up in selection medium (DMEM, 5% FKS, 600 µg/ml Geneticin).

## 2.23 Chromatographic purification of the GP120 mixture

The chromatographic purification takes place according to standard methods (Techniques in HIV Research, Aldovini & Walker, Stockton Press, 1990; S.W. Pyle et al. Purification of 120,00 dalton envelope glycoprotein from culture fluids of human immunodeficiency virus (HIV)-infected H9 cells. AIDS Res Hum Retroviruses 1987, 3:387-400). Cell extracts and cell culture overhangs of GP120-expressing COS cells are used for the isolation of the GP120 mixture. After centrifuging (25,000 g), the lysate is purified as follows:

1.  Gel filtration (Sephadex G-20)
2.  Immunoaffinity chromatography (anti-GP120 antibodies bound to CH-sepharose 4B)
3.  Concentration through protein precipitation
4.  Concentration through dialysis

## 2.24 Preparation of the DNA vaccine

Starting material for the preparation of the DNA vaccine is the mixture of the BstEII-BamHI DNA fragments of the *env* gene which were used for the preparation of the GP120 protein mixture. A eukaryotic expression vector for the HIV-1 GP120 which is approved for DNA vaccinations (cf.

e.g. J.D. Boyer et al., J Infect Dis 1997, 176:1501-1509; J. D. Boyer et al., Nat Med 1997, 3:526-532; M.L. Bagarazzi et al., J Med Primatol 1997, 26:27-33) is modified such that in the *env* gene the cleavage sites for BstEII and BamHI are located at identical points of the *gp120* reading frame. The mutated, variable *env*-gene fragments (BstEII-BamHI) are cloned into such a vector. The modification of the *env* genes in the V2 loop and V3 loop regions and their cloning into the DNA vaccine vector takes place analogously to the cloning steps which are carried out for the preparation of the gp120 mixture.

## Patent Claims

1.  Protein vaccine which comprises a mixture of viral
    protein molecules, **characterized in that** the molecules
    are sequence variants of a single viral protein or of
    part of same, the mixture containing $\geq 10^2$ sequence
    variants, which is obtainable by expression of a
    plasmid-DNA mixture which, because of the variation of
    nucleotide positions, contains randomly distributed
    sequence combinations.

2.  Protein vaccine according to claim 1, **characterized in
    that** the mixture contains $\geq 10^3$ and preferably $\geq 10^4$
    sequence variants.

3.  Protein vaccine according to claim 1 or 2,
    **characterized in that** it comprises a mixture of GP120
    proteins of HIV which in each case differ from each
    other in their amino acid sequence in the region of
    the V2 loop and/or of the V3 loop.

4.  DNA vaccine which codes for a mixture of structurally
    different virus proteins, **characterized in that** the
    vaccine contains a mixture of sequence variants of a
    viral DNA molecule or of part of same, the mixture
    containing $\geq 10^2$ DNA molecules which differ from each
    other in their nucleic acid sequence, where the
    mixture, because of the variation of nucleotide
    positions, contains randomly distributed sequence
    combinations.

AMENDED PAGE

5. DNA vaccine according to claim 4, **characterized in that** it contains a mixture of DNA molecules which code for the sequence variants of a viral protein or a part.

6. DNA vaccine according to claim 4 or 5, **characterized in that** the mixture contains $\geq 10^3$ and preferably $\geq 10^4$ DNA molecules which differ from each other in their nucleic acid sequence.

7. DNA vaccine according to one of claims 4 to 6, **characterized in that** it codes for a mixture of structurally different GP120 proteins of HIV, in which the vaccine contains a mixture of DNA molecules, the nucleic acid sequences of which differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop.

8. DNA vaccine according to claim 7, **characterized in that** it contains a mixture of DNA molecules which differ from each other in their nucleic acid sequence such that they code for a mixture of GP120 proteins which contain amino acid sequences which differ from each other in the V2 loop and/or in the V3 loop.

9. Nucleic acid sequence which is derived from the *env* sequence represented in SEQ ID NO: 1 or a fragment of same, **characterized in that** it is modified such that it contains monovalent restriction sites which make possible the specific exchange of the V2 and V3 regions.

AMENDED PAGE

10. Nucleic acid sequence which is derived from the *env* sequence represented in SEQ ID NO: 1 or a fragment of same, **characterized in that** it is modified such that it contains ten monovalent restriction cleavage sites at intervals of approx. 150 base pairs.

11. Nucleic acid sequence according to claim 9 or 10, **characterized in that** the sequence is modified by the introduction of silent mutations.

12. Nucleic acid sequence according to claims 9 to 11, **characterized in that** it contains the sequence given in SEQ ID NO: 9.

13. Nucleic acid sequence, **characterized in that** it contains the sequence given in SEQ ID NO: 11.

14. Nucleic acid sequence, **characterized in that** it contains the sequence given in SEQ ID NO: 12.

15. Single-stranded nucleic acid sequence which contains the region coding for the V3 loop and/or for the V2 loop of GP120, **characterized in that** in the V3 loop a 247 bp-long BglII-XbaI fragment or a 283 bp-long BglII-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-BclI fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions.

16. Single-stranded nucleic acid sequence which contains the region coding for the V2 loop and/or the region of GP120 coding for the V3 loop, **characterized in that** in the V3 loop a 247 bp-long BglII-XbaI fragment or a 283 bp-long BglII-NheI fragment is exchanged for a modified fragment, and in the V2 loop a 139 bp-long PstI-BclI fragment or a 339 bp-long PstI-EcoRI fragment is exchanged for a modified fragment, the fragment/the fragments in each case containing inosine, a nucleic acid exchange or a mutation at at least 6 positions, the nucleic acid sequence being complementary to the nucleic acid sequence according to claim 14.

17. Nucleic acid sequence according to claim 15 or 16, characterized in that the fragment/the fragments contain(s) inosine, a nucleic acid exchange or a mutation at 9 to 20 positions.

18. Double-stranded DNA which comprises hybrids of the single-stranded nucleic acid sequence according to claim 15 or 17 with the single-stranded nucleic acid sequence according to claim 16 or 17.

19. Nucleic acid mixture which comprises double-stranded DNAs, the nucleic acid sequences of which are derived from the *env* sequence in SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same, **characterized in that** the nucleic acid sequences in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop.

AMENDED PAGE

20.    Nucleic acid mixture according to claim 19,
       **characterized in that** the nucleic acid sequences
       differ from each other such that they code for a
       mixture of proteins which in each case contain amino
       acid sequences which differ from each other in the V2
       loop and/or in the V3 loop.

21.    Nucleic acid mixture according to claim 20,
       **characterized in that** the mixture contains $\geq 10^2$
       sequence variants.

22.    Nucleic acid sequence according to claim 21,
       **characterized in that** the mixture contains $\geq 10^3$ and
       preferably $\geq 10^4$ sequence variants.

23.    Protein mixture which comprises sequence variants of
       the GP120 protein, **characterized in that** it is a
       mixture of proteins which contain amino acid sequences
       which in each case differ from each other in the V2
       loop and/or in the V3 loop, the mixture being
       obtainable by expression of a plasmid DNA mixture
       which, because of the variation of nucleotide
       positions, contains randomly distributed sequence
       combinations.

24.    Protein mixture according to claim 23, **characterized
       in that** the mixture contains $\geq 10^2$ sequence variants.

25.    Protein mixture according to claim 24, **characterized
       in that** the mixture contains $\geq 10^3$ and preferably $\geq 10^4$
       sequence variants.

AMENDED PAGE

26. Plasmid, which contains an inserted double-stranded DNA according to claim 18.

27. Expression vector, **characterized in that** it contains an inserted nucleic acid sequence according to claims 9 to 14.

28. Expression vector according to claim 27, **characterized in that** it contains the sequence given in SEQ ID NO: 10.

29. Expression vector, **characterized in that** it corresponds to DSM 12612.

30. Vector mixture which contains a mixture of plasmids according to claim 26, **characterized in that** the nucleic acid sequences of the plasmids differ in each case from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop, where the mixture of the plasmids, because of the variation of nucleotide positions, contains randomly distributed sequence combinations.

31. Vector mixture according to claim 30, **characterized in that** the mixture contains $\geq 10^2$ plasmids which differ from each other in their nucleic acid sequence.

32. Vector mixture according to claim 31, **characterized in that** the mixture contains $\geq 10^3$ and preferably $\geq 10^4$

AMENDED PAGE

plasmids which differ from each other in their nucleic acid sequence.

33. Vector mixture according to one of claims 30 to 32, **characterized in that** the plasmids can be expressed in *E. coli* as host cell.

34. Vector mixture according to one of claims 30 to 32, **characterized in that** the plasmids can be expressed in eukaryotic cells, preferably in Cos, CHO or BHK cells, as host cells.

35. *E. coli* host cells which are transfected with a vector mixture according to claim 33.

36. Eukaryotic host cells which are transfected with a vector mixture according to claim 34.

37. Eukaryotic host cell according to claim 36, **characterized in that** they are a host cell from the group consisting of Cos, BHK or CHO cells.

38. Process for the preparation of the nucleic acid sequence according to claim 10, **characterized in that** so many silent mutations are introduced into a nucleic acid sequence coding for a viral protein that the nucleic acid sequence thereby obtained contains monovalent restriction sites which make possible the exchange of the V2 and V3 regions.

39. Process for the preparation of the nucleic acid sequence according to claim 10, **characterized in that**

AMENDED PAGE

so many silent mutations are introduced into a nucleic acid sequence coding for a viral protein that the nucleic acid sequence thereby obtained contains ten monovalent restriction sites at intervals of approx. 150 base pairs.

40.   Process according to claim 38 or 39, **characterized in that** the nucleic acid sequence coding for a viral protein is the sequence according SEQ ID NO: 1 or SEQ ID NO: 11 or a fragment of same.

41.   Process for the preparation of the vector mixture according to claims 33 and 34, **characterized in that** plasmids, the nucleic acid sequences of which in each case differ from each other in the region coding for the V2 loop and/or in the region coding for the V3 loop in each case through random distribution of the bases at the varied nucleotide positions, are ligated into a vector which can be expressed in host cells.

42.   Process according to claim 41, characterized in that the host cells are *E. coli*, Cos, CHO or BHK cells.

43.   Process for the preparation of the host cells according to claim 35 or 36, **characterized in that** the host cells are transformed with a vector mixture according to claims 30 to 32.

44.   Process for the preparation of a protein vaccine according to one of claims 1 to 3, **characterized in that** the host cells are cultivated according to one of claims 35 to 37 under conditions which allow the

AMENDED PAGE

expression of the mixture of viral protein sequence variants.

45. Process for the preparation of a DNA vaccine according to one of claims 4 to 8, **characterized in that** the process is carried out according to claim 41 or 42, the plasmids according to the invention being ligated into a vector which can be expressed in host cells of the organism to be vaccinated.

46. Use of a mixture of structurally different viral proteins which are sequence variants of a viral protein or of part of same, for the preparation of a vaccine for the prevention and/or therapy of a virus infection in humans.

47. Use of a protein mixture according to one of claims 23 to 25 for the preparation of a vaccine for the prevention and/or therapy of a HIV infection in humans.

48. Use of a mixture of DNA molecules which code for sequence variants of a viral protein or of part of same, for the preparation of a vaccine for the prevention and/or therapy of a virus infection in humans.

49. Use of a nucleic acid mixture according to claims 19 to 22 for the preparation of a vaccine for the prevention and/or therapy of a virus infection in humans.

AMENDED PAGE

50. Use of the nucleic acid mixture according to one of claims 19 to 22 for the preparation of a vector mixture according to one of claims 30 to 32 which can be expressed in host cells , the host cells being selected from the group consisting of *E. coli*, Cos, CHO and BHK cells.

51. Use of the vector mixture according to one of claims 30 to 32 for the expression of a protein mixture according to one of claims 23 to 25.

52. Use of the host cell according to one of claims 35 to 37 for the preparation of a protein mixture according to one of claims 23 to 25.

53. Pharmaceutical composition for the prevention and/or therapy of a virus infection, **characterized in that** it comprises a protein mixture and a nucleic acid mixture, the protein mixture comprising sequence variants of a viral protein or of part of same, and the nucleic acid mixture comprising DNA molecules which code for sequence variants of a viral protein or of part of same.

54. Pharmaceutical composition according to claim 53, **characterized in that** it comprises a protein mixture according to claims 23 to 25 and a nucleic acid mixture according to claims 19 to 22.

AMENDED PAGE

# Abstract

The present invention relates to a pharmaceutical
composition or a vaccine which comprises a mixture of viral
protein molecules which are sequence variants of a single
viral protein or of part of same. The invention furthermore
relates inter alia to a DNA vaccine which codes for a
mixture of structurally different virus proteins, the
vaccine containing a mixture of sequence variants of a
viral DNA molecule or of part of same which code for
sequence variants of a viral protein or part. According to
a preferred version of the invention, the viral proteins
are sequence variants of the GP120 protein of the human
immunodeficiency virus (HIV) which differ from each another
in their amino acid sequence in the region of the V2 loop
and/or of the V3 loop. The invention furthermore relates to
the preparation of the virus vaccines including the
intermediate stages or constructs, preparation processes
and uses connected with them.

**Fig. 1**

V3 loop sequence data of HIV-1 patient isolates (PI)

**Fig. 2a**

Amino acid sequences of the NL4-3 V2 loop and NL4-3 V3 loop. The regions which are preferably to be varied are underlined.

**Fig. 2b**

Variation of the V3 loop. Data from the Los Alamos database.

**Fig. 4**

Schematic representation of the process for the preparation of the mixture of gp 120-expressing plasmid vectors.

1.) Preparation of the degenerated DNA fragments for e.g. the V3 loop
   a.) Synthesis of single-stranded DNA
   b.) Hybridization of two complementary oligonucleotides.
2.) Cloning of the V3 loop DNA fragments into pUC18 delta *env*
3.) Cloning of the *env* gene into the gp 120 expression vector pBSCenvATG

   V3 loop fragments with degenerated sequence e.g.:

Fig. 5

Heterogeneity of the vaccine using the V3 loop as an
example

V3 loop sequence:

Degeneration at protein level:

variants

Degenerated DNA sequence of the respective variable amino
acid positions:

variants

Variability at protein level derived from the degenerated
DNA sequence:

**RULE 63 (37 C.F.R. 1.63)**
**INVENTORS DECLARATION FOR PATENT APPLICATION**
**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

As a below named inventor, I hereby declare that my residence, mailing address and citizenship are as stated below next to my name, and I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

VIRAL VACCINE

the specification of which (**check** applicable **box(s)**).

☐ is attached hereto
☐ was filed on _____ as U.S. Application Serial No. _____ (Atty Dkt. No. 35-213)
☒ was filed as PCT International application No. _____ PCT/EP99/09759 ___ on ___ 3 December 1999
and (if applicable to U.S. or PCT application) was amended on _____ 2 November 2000

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose to the Patent Office all information known to me to be material to patentability as defined in 37 C.F.R. 1.56. I hereby claim foreign priority benefits under 35 U.S.C. 119/365 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed or, if no priority is claimed, before the filing date of this application:
Priority Foreign Application(s):

| Application Number | Country | Day/Month/Year Filed |
|---|---|---|
| 199 07 485.2 | GERMANY | 12 February 1999 |

I hereby claim the benefit under 35 U.S.C §119(e) of any United States provisional application(s) listed below.

| Application Number | Date/Month/Year Filed |
|---|---|
| | |

I hereby claim the benefit under 35 U.S.C. 120/365 of all prior United States and PCT international applications listed above or below.

**Prior U.S./PCT Application(s):**

| Application Serial No. | Day/Month/Year Filed | Status: patented pending, abandoned |
|---|---|---|
| PCT/EP99/09759 | 3 December 1999 | |

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon. And on behalf of the owner(s) hereof, I hereby appoint **NIXON & VANDERHYE P.C., 1100 North Glebe Rd., 8th Floor, Arlington, VA 22201-4714, telephone number (703) 816-4000 (to whom all communications are to be directed)**, and the following attorneys thereof (of the same address) individually and collectively owner's/owners' attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith and with the resulting patent: Larry S. Nixon, 25640; Arthur R. Crawford, 25327; James T. Hosmer, 30184; Robert W. Faris, 31352; Richard G. Besha, 22770; Mark E. Nusbaum, 32348; Michael J. Keenan, 32106; Bryan H. Davidson, 30251; Stanley C. Spooner, 27393; Leonard C. Mitchard, 29009; Duane M. Byers, 33363; Jeffry H. Nelson, 30481; John R. Lastova, 33149; H. Warren Burnam, Jr. 29366; Mary J. Wilson, 32955; J. Scott Davidson, 33489; Alan M. Kagen, 36178; Robert A. Molan, 29834; B. J. Sadoff, 36663; James D. Berquist, 34776; Updeep S. Gill, 37334; Michael J. Shea, 34725; Donald L. Jackson, 41090; Michelle N. Lester, 32331; Frank P. Presta, 19828; Joseph S Presta, 35329; Joseph A. Rhoa, 37515; Raymond Y. Mah, 41426; Chris Comuntzis, 31097. I also authorize Nixon & Vanderhye to delete any attorney names/numbers no longer with the firm and to act and rely solely on instructions directly communicated from the person, assignee, attorney, firm, or other organization sending instructions to Nixon & Vanderhye on behalf of the owner(s).

1.  Inventor's Signature: _____ Date: 19-89-2001
    Inventor: Michael _____ MI _____ SCHREIBER _____ Germany
              (first)           (last)      (citizenship)
    Residence: (city) Hamburg _____ (state/country) Germany
    Mailing Address: Heidberg 35, Hamburg, Germany
    (Zip Code) D-22301

# United States Patent & Trademark Office
## Office of Initial Patent Examination -- Scanning Division
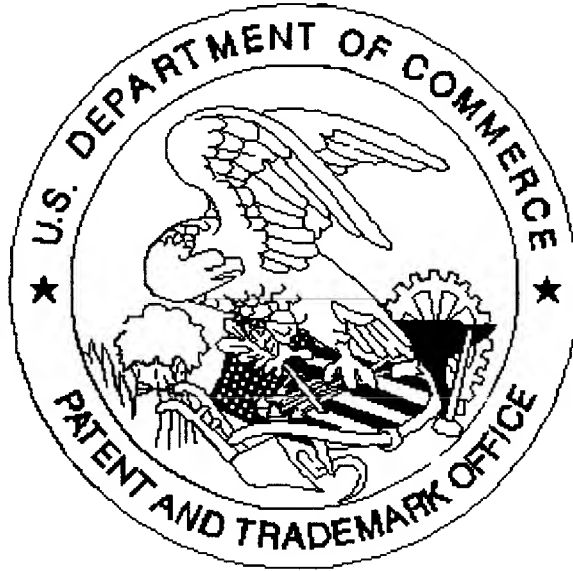
SCANNED, # 14

Application deficiencies found during scanning:

☑ Page(s)_____ of ___Drawings_____ were not present
for scanning.                              (Document title)

☐ Page(s)_____ of_____were not present
for scanning.                              (Document title)

☐ *Scanned copy is best available.*